



QY 2988 CCGCTGCTGAGGGGACCCGGTGCAGTACCGCATGATGACCATCTCTCCATGGCGCA 3047  
Db |||||  
1449 GACGTTGGCCGGGCCACCGCTGGAGGAGAGATGTCGCTCAGCTCTTCATGAAGG 1508  
QY 3048 CGGCT 3053  
Db |||||  
1509 CGGCT 1514

## RESULT 2

US-09-158-767-10  
; Sequence 10, Application US/09158767A  
; Patent No. 6180363  
; GENERAL INFORMATION:  
; APPLICANT: Bataud, Yannick  
; APPLICANT: Durst, Francis  
; APPLICANT: Werck-Reichhart, Daniele  
; TITLE OF INVENTION: RECORDING OF DNA SEQUENCES PERMITTING  
; FILE REFERENCE: A32000  
; CURRENT APPLICATION NUMBER: US/09/158,767A  
; PRIOR FILING DATE: 1998-09-23  
; EARLIER APPLICATION NUMBER: FR 97-12094  
; PRIOR FILING DATE: 1997-09-24  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 10  
; LENGTH: 2181  
; TYPE: DNA  
; ORGANISM: Triticum aestivum  
US-09-158-767-10

Query Match 3.1%; Score 119.6; DB 3; Length 2181;  
Best Local Similarity 59.8%; Pred. No. 3.4e-20;  
Matches 219; Conservative 0; Mismatches 144; Indels 3; Gaps 1;

QY 2691 CGCGCAGGACCCCAAGGGATCCTGGAGGACGACGTCCTCCGCGGAGCGGAGCGAGTGGAG 2750  
Db |||||  
1260 CCGGAGGACTCAAGCAGCTGTCGGGAGGACTACTCCCCGAGCGGACCTTCGTGGC 1319  
QY 2751 GCGCGCGGGATGTCAGTACGTCGCTGATGGGCGGATGAGTCAACTGGGG 2810  
Db |||||  
1320 GCGCGGCTGTCGTCACCTACTCCATATACCTCGCGGGCGCATGAAGGGGTGGGG 1379  
QY 2811 CCGCGAGCGGAGCTTCGCGCGGAGCGGTGGATCAAGAGATGGCGC---GTTCCG 2867  
Db |||||  
1380 GGAGGACTGCTGAGTTCGCGCGGAGCGATGGCTGTCGCGCGAGCGGACCAAGTTGGA 1439  
QY 2868 CAACGCTGCGCGTTCAAGTTCACGCGGTTCCAGGCGGGGCGGAGATCTGCTGGGCAA 2927  
Db |||||  
1440 GCAGCAGGACTCGTACAAGTTCGTGGCGTTCAACGCGGCGGAGGTGTGCTGGGCAA 1499  
QY 2928 GGAATCGGCGTACCTCGATGAGATGGCGTGGCCATCTCTTTCGCTTTCACAGCTT 2987  
Db |||||  
1500 GGAATCGGCGTACCTCGATGAGATGGCGTGGCCATCTCTTTCGCTTTCACAGCTT 1559  
QY 2988 CCGGCTGCTGGAGGGGACCCGCTGCGATGACCATCTCTTCATGGCGCA 3047  
Db |||||  
1560 GACCGTGGCGCGGGCCACCGCTGGAGGAGATGTCGTCACGCTCTTCATGAAGG 1619  
QY 3048 CGGCT 3053  
Db |||||  
1620 CGGCT 1625

## RESULT 3

US-09-410-551B-1  
; Sequence 1, Application US/09410551B  
; Patent No. 6503737  
; GENERAL INFORMATION:  
; APPLICANT: KOSAN BIOSCIENCES, Inc.  
; APPLICANT: REEVES, CHRISTOPHER

; APPLICANT: CHU, DANIEL  
; APPLICANT: KHOSLA, CHAITAN  
; APPLICANT: SANTI, DANIEL  
; APPLICANT: WU, KAI  
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA  
; FILE REFERENCE: 30062-20026.00  
; CURRENT APPLICATION NUMBER: US/09/410,551B  
; PRIOR FILING DATE: 1999-10-01  
; PRIOR APPLICATION NUMBER: US 60/139,650  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/123,810  
; PRIOR FILING DATE: 1999-03-11  
; PRIOR APPLICATION NUMBER: US 60/102,748  
; PRIOR FILING DATE: 1998-10-02  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 77536  
; TYPE: DNA  
; ORGANISM: Streptomyces hygroscopicus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (52275)....(71465)  
US-09-410-551B-1

Query Match 1.7%; Score 65.4; DB 4; Length 77536;  
Best Local Similarity 51.9%; Pred. No. 1.5e-05;  
Matches 175; Conservative 0; Mismatches 156; Indels 6; Gaps 1;

QY 2232 ATCGAGTAGGCGAGCGCGGACGACGCGCGGCTTCGGGAGCAGACAGCTCCGG 2291  
Db |||||  
56223 ACCGGCCCGCGGTGACGACCTGTCCCGCGGACGCGGTGTCGGCCTGACCCGGGC 56282  
QY 2292 GACGTGGTCTCAACTTCGTGATCGCGGCGGACACGCGGACGACGCTGTCTGG 2351  
Db |||||  
56283 GGCATCGGCCCGAGCGGCTTACGACCGCGCTGGCTGCGCGGATCCCGACGCGCTGC 56342  
QY 2352 TTCAGCGACATGGCCATGTCCTCCACCGGACGTCGGCGGAGTGGCGCGAGTGTGC 2411  
Db |||||  
56343 AGCTTACACGCGCGGCTCGTCCGATCGTGTTCGGGACCGCGTGTGACGCGCTGTC 56402  
QY 2412 GGTTCGAGGCGGCGCGCGCGGAGGAGCGCTGCGCTGCGCCCTGCGCGCGCGCT 2471  
Db |||||  
56403 GACCTCGGACACTGCGCGCGGCGAGAGTCTCTGTCACGCGGCCACCGCGCGGTGTC 56462  
QY 2472 GACGCGGACGACGAGCGTTCGCGCGCGGCGGCGAGTTCGCGGGGCTCTCACCTAC 2531  
Db |||||  
56463 GGCATGGCGCGGACAGATCGCGCGCACCTGG-----GGCGGAGTCTACGCCACC 56516  
QY 2532 GACAGCTCGGCAAGCTGCTTACCTCAAGCTGCG 2568  
Db |||||  
56517 GCAGTACCGGCAAGCAGCAGCTCTGCGCGCGCGCG 56553

## RESULT 4

US-09-452-239-41  
; Sequence 41, Application US/09452239  
; Patent No. 6465229  
; GENERAL INFORMATION:  
; APPLICANT: Rafalski, Antoni J.  
; APPLICANT: Pader, Gary M.  
; APPLICANT: Cahoon, Rebecca E.  
; TITLE OF INVENTION: Plant Caffeoyl-CoA O-Methyltransferase  
; FILE REFERENCE: BB1284 US NA  
; CURRENT APPLICATION NUMBER: US/09/452,239  
; CURRENT FILING DATE: 1999-12-01  
; EARLIER APPLICATION NUMBER: 60/110,594  
; EARLIER FILING DATE: 1998-December-02  
; NUMBER OF SEQ ID NOS: 50  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 41  
; LENGTH: 1078

```

; TYPE: DNA
; ORGANISM: Triticum aestivum
US-09-452-239-41

Query Match      1.6%; Score 62.8; DB 4; Length 1078;
Best Local Similarity 48.1%; Pred. No. 5.9e-06;
Matches 178; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

QY 2252 CGACGACGCGCGCTTCGGGAGACGACAGAGCCTTCGGGACGTGGTCTCAACTTCGT 2311
DB 423 CCGCACCGGCTCGGCATCCCGACGACGCGCACTTCGGCCATGACATCAACCGGGA 482
QY 2312 GATCCCGGGGGGACACAGACGCGACGAGCTGTCGTGTTTACGACATGGCCATGTC 2371
DB 483 GAACTACGAGCTGGGGCTGCGGTGATCGAGAGCGCGGTGGGCGCAAGATCGACTT 542
QY 2372 CCACCGGAGTGGCCGAGAGCTCGGCGGAGCTGTCGCTTCGAGGCGGAGCGCGC 2431
DB 543 CCGGAGGCGCGCGCTGCGGCTGCGGCGGCGCTGAGCGCGAGCAAGCGGCAACCA 602
QY 2432 CGCGGAGGAGGCGTTCGCTGCTGCTGCGGCGGCGCTGAGCGCGAGCAAGCGGCTT 2491
DB 603 CACTTCGACTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 662
QY 2492 CGCGCGCGCGTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 2551
DB 663 CACTGAGAGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 722
QY 2552 CTACTTCACGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 2611
DB 723 CTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 782
QY 2612 CGCGCGCGGAC 2621
DB 783 GTCCTCGAC 792

RESULT 5
US-09-249-585A-4/c
; Sequence 4, Application US/09249585A
; Patent No. 6417002
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert
; TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISOMES
; FILE REFERENCE: 0867/0D905
; CURRENT APPLICATION NUMBER: US/09/249,585A
; CURRENT FILING DATE: 1999-02-11
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 1926
; TYPE: DNA
; ORGANISM: Epstein Barr Virus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1926)
; OTHER INFORMATION: template strand of EBNA-1 DNA
US-09-249-585A-4

Query Match      1.5%; Score 59.8; DB 4; Length 1926;
Best Local Similarity 43.9%; Pred. No. 4.7e-05;
Matches 306; Conservative 0; Mismatches 387; Indels 4; Gaps 1;

QY 2241 GCGAGGCGCGGACGACGCGCGCGCTTCGGGAGACGACGAGCCTTCGGGACGTGGT 2300
DB 952 GACGGGAGGAGACGAGGACGCGGAGGACGCGGAGGACGAGGACGCGGAGGACGCGG 893
QY 2301 CTCAACTTCGTATCGCGCGGCGGACACGACGCGGACGACGCTGTCGTGTTTACGAC 2360
DB 892 GACGAGGACGCGGAGGACGCGGAGGACGAGGACGCGGAGGACGCGGAGGACGAGGAC 833
QY 2361 ATGGCCATGTCCTCCACCGCGAGCTGCGCGCGGAGCTGCGCGCGGAGCTGTCGCTTCGAG 2420

; TYPE: DNA
; ORGANISM: Triticum aestivum
US-09-452-239-41

Query Match      1.6%; Score 62.8; DB 4; Length 1078;
Best Local Similarity 48.1%; Pred. No. 5.9e-06;
Matches 178; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

QY 2252 CGACGACGCGCGCTTCGGGAGACGACAGAGCCTTCGGGACGTGGTCTCAACTTCGT 2311
DB 423 CCGCACCGGCTCGGCATCCCGACGACGCGCACTTCGGCCATGACATCAACCGGGA 482
QY 2312 GATCCCGGGGGGACACAGACGCGACGAGCTGTCGTGTTTACGACATGGCCATGTC 2371
DB 483 GAACTACGAGCTGGGGCTGCGGTGATCGAGAGCGCGGTGGGCGCAAGATCGACTT 542
QY 2372 CCACCGGAGTGGCCGAGAGCTCGGCGGAGCTGTCGCTTCGAGGCGGAGCGCGC 2431
DB 543 CCGGAGGCGCGCGCTGCGGCTGCGGCGGCGCTGAGCGCGAGCAAGCGGCAACCA 602
QY 2432 CGCGGAGGAGGCGTTCGCTGCTGCTGCGGCGGCGCTGAGCGCGAGCAAGCGGCTT 2491
DB 603 CACTTCGACTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 662
QY 2492 CGCGCGCGCGTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 2551
DB 663 CACTGAGAGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 722
QY 2552 CTACTTCACGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 2611
DB 723 CTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 782
QY 2612 CGCGCGCGGAC 2621
DB 783 GTCCTCGAC 792

RESULT 6
US-09-130-114-2/c
; Sequence 2, Application US/09130114
; Patent No. 5978807
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert A.
; APPLICANT: Damaj, Bassam B.
; APPLICANT: Robbins, Alan K.
; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
; FILE REFERENCE: 0867/1D903US1
; CURRENT APPLICATION NUMBER: US/09/130,114
; CURRENT FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1931
; TYPE: DNA
; ORGANISM: EBNA
US-09-130-114-2

Query Match      1.5%; Score 59.8; DB 2; Length 1931;
Best Local Similarity 43.9%; Pred. No. 4.7e-05;
Matches 306; Conservative 0; Mismatches 387; Indels 4; Gaps 1;

QY 2241 GCGAGGCGCGGACGACGCGCGCGCTTCGGGAGACGACGAGCCTTCGGGACGTGGT 2300
DB 952 GACGGGAGGAGACGAGGACGCGGAGGACGCGGAGGACGAGGACGCGGAGGACGCGG 893
QY 2301 CTCAACTTCGTATCGCGCGGCGGACACGACGCGGACGACGCTGTCGTGTTTACGAC 2360
DB 892 GACGAGGACGCGGAGGACGCGGAGGACGAGGACGCGGAGGACGCGGAGGACGAGGAC 833
QY 2361 ATGGCCATGTCCTCCACCGCGAGCTGCGCGCGGAGCTGCGCGCGGAGCTGTCGCTTCGAG 2420
```



LOCATION: 46851...47891  
 OTHER INFORMATION: /product= "Sorm"  
 OTHER INFORMATION: /note= "The protein encoded by the sorm gene is highly homologous to the methyltransferase from Streptomyces hygroscopicus that is involved in the synthesis of the polyketide rapamycin."  
 US-08-764-233A-1

Query Match  
 Best Local Similarity 50.5%; Pred. No. 0.00078;  
 Matches 141; Conservative 0; Mismatches 138; Indels 0; Gaps 0;  
 2243 CGAGCCGGCGACGACGCGCGCGCTTCGGGGAGCAAGAGCTCCGGGAGTGGTCT 2302  
 Db 8502 CGAGCCGGCGATGGCGGGGCTCTTCGAGGATGACGTGAGCGTTGGTCCGGAGATC 8443  
 2303 CAACCTTGTGATCCCGCGCGGGACACGACGCGGACGACGCTGCTGCTTACGACAT 2362  
 Db 8442 CGAAGCGGAGACCCCGCGCGGGGAGCCCATCCGCTGCGGACCCCAAGGCTCG 8383  
 2363 GGCCATGTCCACCCGCGACGTGGCGGAGAGCTGCGCGGAGCTGCGGGTTCGAGGC 2422  
 Db 8382 TGGCATCGACGCGGACGCGGAGCGCATCCACGCGATCAGGGGATTGCGTGGGTGG 8323  
 2423 GGAGCGCGCGGAGGAGGCGCTGCGCTGCTGCGCGCGGCTGACGCGGACGA 2482  
 Db 8322 TGTGAGGCTGGGGGCGAGGCGCTGTGTCAGGAGGAGGACGATCTTGGCGAGCCCG 8263  
 2483 CAAGGCGTTGCGCGCGCGCTGGCGGAGTTCGCGGGCT 2521  
 Db 8262 CGAGGCGCGCGCGCTCGAGTGGGCCAACGTTGGTCT 8224

## RESULT 8

US-09-197-649-7  
 ; Sequence 7, Application US/09197649  
 ; Patent No. 6194550  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gold, Larry  
 ; APPLICANT: Turk, Craig  
 ; APPLICANT: Pribrnow, David  
 ; APPLICANT: Smith, Jonathan D.  
 ; TITLE OF INVENTION: Systematic Polypeptide Evolution by Reverse Translation  
 ; FILE REFERENCE: NEX02/C1-CON  
 ; CURRENT APPLICATION NUMBER: US/09/197,649  
 ; CURRENT FILING DATE: 1998-11-23  
 ; EARLIER APPLICATION NUMBER: 07/829,461  
 ; EARLIER FILING DATE: 1992-01-31  
 ; EARLIER APPLICATION NUMBER: 07/739,055  
 ; EARLIER FILING DATE: 1991-08-01  
 ; EARLIER APPLICATION NUMBER: 07/561,968  
 ; EARLIER FILING DATE: 1990-08-02  
 ; NUMBER OF SEQ ID NOS: 26  
 ; SOFTWARE: Patent In Ver. 2.0  
 ; SEQ ID NO 7  
 ; LENGTH: 390  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: Sequence  
 ; OTHER INFORMATION: having a 120 repeat of ACG flanked by fixed  
 ; OTHER INFORMATION: fragments having NcoI restriction sites.  
 US-09-197-649-7

Query Match  
 Best Local Similarity 50.3%; Pred. No. 5.4e-05;  
 Matches 195; Conservative 0; Mismatches 190; Indels 3; Gaps 2;  
 2227 GGTTCATCAGCTAGCGGAGGCGGCGACGACGCGCGGCTTCGGGAGCAGAGCC 2286  
 Db 1 GGGCCATGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 60  
 2287 TCCGGGACGTGTGTCTCACTTCGTGATCGCGGGCGGGAC-ACGAGCGCGACGCGTGT 2345

Db 61 ACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 120  
 QY 2346 TCGTGGTTTCGACACATGGCCATGTC--CAACCCGGAGTGGCGGAGAGTGGCCCGG 2403  
 Db 121 ACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 180  
 QY 2404 AGTGTGCGGCTTCAGGCGGAGCGCGCGGCGGAGGCGGCTCGCTCGTGCCTGCG 2463  
 Db 181 ACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 240  
 QY 2464 GCGCGCTGAGCGCAGCAGCAAGCGCTTTCGCGCGCGCGTGGCGCACTTTCGCGGCGCTCC 2523  
 Db 241 ACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 300  
 QY 2524 TCACCTACGACAGCGCTCGGCAAGCTGCTTACCTCCAGCGCTGCTCAGCAGAGCGTCC 2583  
 Db 301 ACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 360  
 QY 2584 GCCTGTACCCCGCGCTCCCTCAGGTGAG 2611  
 Db 361 ACAGCAGCAGCAGTGGTTGCGATGCTCAG 388

## RESULT 9

US-09-266-965-98  
 ; Sequence 98, Application US/09266965  
 ; Patent No. 6495348  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Sherman, D  
 ; APPLICANT: Mao, Y  
 ; APPLICANT: Varoglu, M  
 ; APPLICANT: He, M  
 ; APPLICANT: Sheldon, P  
 ; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster  
 ; FILE REFERENCE: 600.456US1  
 ; CURRENT APPLICATION NUMBER: US/09/266,965  
 ; CURRENT FILING DATE: 1999-03-12  
 ; EARLIER APPLICATION NUMBER: US 08/624,447  
 ; EARLIER FILING DATE: 1996-08-19  
 ; EARLIER APPLICATION NUMBER: PCT/US94/11279  
 ; EARLIER FILING DATE: 1994-10-06  
 ; EARLIER APPLICATION NUMBER: US 08/133,963  
 ; EARLIER FILING DATE: 1993-10-07  
 ; NUMBER OF SEQ ID NOS: 145  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 98  
 ; LENGTH: 819  
 ; TYPE: DNA  
 ; ORGANISM: Streptomyces lavendulae  
 US-09-266-965-98

Query Match  
 Best Local Similarity 46.4%; Pred. No. 0.00012;  
 Matches 255; Conservative 0; Mismatches 291; Indels 3; Gaps 2;  
 QY 2329 CGACGGCGACGACGCTGTGTTTACGACATGCCATGTCCACCGGACGTGGCG 2388  
 Db 6 CTTGTCGACGACACGACGCGGGGCGACGCGGTGATCACTGATCGCGTGCACAGGCA 65  
 QY 2389 AGAAGCTGGCGCGAGTGTGCGGTTTCGAGGCGGAGCGCGCGCGGAGGAGGCGTCC 2448  
 Db 66 GATGAGCGGCTTCTTCCGCTGCGGTGGTGGGACCGC-TGCGGGGGGGCGCGCG 124  
 QY 2449 CGCTGCTCCCTGGCGCGCGCTGACCGGACGACAGCGTTCGCCGCGCGCGTGGCG 2508  
 Db 125 CCGGAACCGCGGGCGCGCGCGCTCGCGCCCCGCTGCTCTCTCTCGACGACGACG 184  
 QY 2509 AGTTTCGCGGCTCTCTACCTACGACAGCTCGGCAAGCTGCTTACCTCCAGCGCTCG 2568  
 Db 185 TCCTGGTGGCGCGGCTTCTTCGCGGACAGCGCGCGCGCGGACCGGACGCGCTTCA 244  
 QY 2569 TCACGAGACGCTCGCGCTGTACCCCGCGCTCCCTCAGGTGAGCGCGCGCGACGCG--G 2626

Db 245 CCACGCGCGCTCGCGGAACTCCACCGCGCGCGTTCCTCGCGCTGTGAGAAAG 304  
QY 2627 ACCTCCGCTCAGAGCACAGCATGCTGAGTGTGACCTGATGCAATGACATGCACTTG 2686  
Db 305 CCGCCCGGACCGAGTCCGCGCGCGCGCGCTCGAAACCGCTGCCCGCGCGCT 364  
QY 2687 CGCGCGCGGAGGACCCCAAGGGGATCTTGAGGAGCAAGCTGTGCGGAGCGGACGAAGG 2746  
Db 365 CCAGCGCGCCCAACCGCACCGCGCGCTCGTCCCAACCGCTTGAGCGCGCGTGGAGG 424  
QY 2747 TGAGGCGCGCGGAGTGTGACGTACCTGCTTACTGATGGCGCGGATGAGTCAACT 2806  
Db 425 CCATGCGCGCGGCTCCCTGCGGAGCTGCGCCCTGGCTCGGCTTCATCGCGCGGAA 484  
QY 2807 GGGGCGCGGACCGCGGAGTTCGCGCGGAGCGGTGATCAACGAGGATGCGGCTTCC 2866  
Db 485 CGGCCCTCGACAAAGCGCGATGGAGCATACCGCGGATTCGACGAGGAGTTCGGGCTCA 544  
QY 2867 GCAACGCGT 2875  
Db 545 CTGCGGGT 553

## RESULT 10

US-09-452-239-35

; Sequence 35, Application US/09452239

; Patent No. 6465229

; GENERAL INFORMATION:

; APPLICANT: Rafalski, Antoni J.

; APPLICANT: Fader, Gary M.

; APPLICANT: Cahoon, Rebecca E.

; TITLE OF INVENTION: Plant Caffeoyl-CoA O-Methyltransferase

; FILE REFERENCE: B01284 US NA

; CURRENT APPLICATION NUMBER: US/09/452,239

; EARLIER FILING DATE: 1999-12-01

; EARLIER FILING DATE: 1998-December-02

; NUMBER OF SEQ ID NOS: 50

; SOFTWARE: Microsoft Office 97

; SEQ ID NO 35

; LENGTH: 1018

; TYPE: DNA

; ORGANISM: Triticum aestivum

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (817)

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (826)

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (874)

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (891)

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (924)

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (934)

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (961)

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (970) .. (971)

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (1012)

; US-09-452-239-35

Query Match 1.5%; Score 57.4; DB 4; Length 1018;  
Best Local Similarity 49.2%; Pred. No. 0.00013;  
Matches 151; Conservative 0; Mismatches 156; Indels 0; Gaps 0;  
QY 2252 CGACGACGCGCGCGCTTCGGGAGCAGCAAGAGCTCCGCGGACGTGCTCAACTTGGT 2311  
Db 400 CGCCACCGCGCTCGGCATCCCGACGACGCGGACCATTTGGCCATGGACATCAACCGCA 459  
QY 2312 GATCCCGCGCGGACACGACGCGGACGAGCTGTGCTGTTCAGGCACATGCGCATGTC 2371  
Db 460 GAACTACGAGCTGGGGTGCCTGTCATGAGAAGCGCGCGTGGCGCAAAAGATCGACTT 519  
QY 2372 CCACCGGACGCTGGCGGAGAGCTCGCGCGGAGCTGTGCGGTTCGAGGCGGAGCGGC 2431  
Db 520 CGCGGAGCGCGCGGCTGCGCGCTCTGAGACGCTGTGGAGGACGAGGCGCAACCGG 579  
QY 2432 GGGGAGGAGGCGCTGCGCTGTGTCCTTGGCGCGCGCTGACGCGGACGCAAGCGCTT 2491  
Db 580 GACCTTCGACTTGGTTCGTTGGAGCGCGCAAGGACAACCTACCTCACTACCGAGCG 639  
QY 2492 CGCGCGCGCGTGGCGGAGTTCGCGGCGCTCTCACTACGACAGCTTCGCGAAGCTGGT 2551  
Db 640 CCTCATGAAGCTCGTCAAGGTGCGCGCGCTCTCGGTACGACAACACCTCTGGAACGG 699  
QY 2552 CTACCTC 2558  
Db 700 CTCGCTC 706

## RESULT 11

US-09-266-965-9

; Sequence 9, Application US/09266965

; Patent No. 6495348

; GENERAL INFORMATION:

; APPLICANT: Sherman, D

; APPLICANT: Mao, Y

; APPLICANT: Varoglu, M

; APPLICANT: He, M

; APPLICANT: Sheldon, P

; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster

; FILE REFERENCE: 600.456US1

; CURRENT APPLICATION NUMBER: US/09/266,965

; EARLIER FILING DATE: 1999-03-12

; EARLIER APPLICATION NUMBER: US 08/624,447

; EARLIER FILING DATE: 1996-08-19

; EARLIER APPLICATION NUMBER: PCT/US94/11279

; EARLIER FILING DATE: 1994-10-06

; EARLIER APPLICATION NUMBER: US 08/133,963

; EARLIER FILING DATE: 1993-10-07

; NUMBER OF SEQ ID NOS: 145

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 9

; LENGTH: 3765

; TYPE: DNA

; ORGANISM: Streptomyces lavendulae

; US-09-266-965-9

Query Match 1.5%; Score 57.4; DB 4; Length 3765;  
Best Local Similarity 46.4%; Pred. No. 0.00028;  
Matches 255; Conservative 0; Mismatches 291; Indels 3; Gaps 2;  
QY 2329 CGACGCGGACGACGCTGTGCTGTTTCAGGCACATGCCATGCCCATGCCGCGCGCGGCGCG 2388  
Db 1884 CGTCGTGACGACAACGACGCGGCGGCGACGCCGCTGATCACTGATGCCGCTGACAGCGA 1943  
QY 2389 AGAAGCTGCGCGCGAGTGTGCGGTTTCGAGCGCGGAGCGCGCGCGGAGGCGGCTCG 2448  
Db 1944 GATGAGCGGCTTCTCCCGCTGCGCGTGTGCGGCGGACCGG-TGCGGGGCGGCGCGCG 2002  
QY 2449 CGTCGTGCTTCGCGCGGCGGCTGACGCGGACGCAAGAGCGTTCCCGCGCGCTGGCGC 2508  
Db 2003 CCCGGAACGCGCGGCGCGCGCGCTCGCGCGCGCTGCTGCTTCTCTGACGACGACG 2062

QY 2509 AGTTCCGGGCTCTCACTACGACAGCCTCGGCAAGCTGGTCTACCTCCAGCCTGG 2568  
Db 2063 TCTGTGGGGCCCGGCTTCTCTCGCGCACACGCCGCGCGCGGAAACCGGACGCTTCA 2122  
QY 2569 TCACCGAGAGCTCCGCTGTACCCCGCGCTCTCTAGGTGAGCGCGCGACACGC--G 2626  
Db 2123 CCACGCGCGCTGCGCGAACTCCCAACCGCGCGGCTCTCTCGCGCTGTGAGAAGG 2182  
QY 2627 ACTCCGGTCCAGAGCACAGCATGAGTGGACCTGAATGCAATGCAATGCAATGCACTTG 2686  
Db 2183 CCGCCCGAGAGTCCGCGCGCGCGCGCGGACTCGAAACCGCTGCCCGCGCGCT 2242  
QY 2687 CCGCGCGAGACCCCAAGGGATCTCGAGAGACAGCTGTCTCGCGAGCGGACGAAGG 2746  
Db 2243 CCGAGCGCGCGCAACCGCACCGCGCGCTCTGCGCAACCGCTGAGCGCGCGTGGAGG 2302  
QY 2747 TGAGGCGCGCGGATGAGTGGAGTCTGAGTGGACCTGAATGCAATGCAATGCACTTG 2806  
Db 2303 CCATGCGCGCGCTCTCGCGAGCTGCGCGCGCTGCGCGCGCTCTGCGCGCGAAACA 2362  
QY 2807 GGGGCGCGACCGCGAGCTTCCGCGCGAGCGGTGGATCAACGAGGATGCGGCTTC 2866  
Db 2363 CCGCCCTCGACAAGCGCGATCGGAGCATACCGCGGATTCGACGAGGATTCGGGCTCA 2422  
QY 2867 GCAACGCGT 2875  
Db 2423 CTTGGGGGT 2431

## RESULT 12

US-09-266-965-76/c

; Sequence 76, Application US/09266965

; Patent No. 6495348

; GENERAL INFORMATION:

; APPLICANT: Sherman, D

; APPLICANT: Mao, Y

; APPLICANT: Varoglu, M

; APPLICANT: He, M

; APPLICANT: Sheldon, P

; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster

; FILE REFERENCE: 600.456U51

; CURRENT APPLICATION NUMBER: US/09/266,965

; EARLIER APPLICATION NUMBER: US/08/624,447

; EARLIER FILING DATE: 1996-08-19

; EARLIER FILING DATE: 1994-10-06

; EARLIER FILING DATE: 1993-10-07

; NUMBER OF SEQ ID NOS: 145

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 76

; LENGTH: 53500

; TYPE: DNA

; ORGANISM: Streptomyces lavendulae

US-09-266-965-76

## Query Match

Best Local Similarity 1.5%; Score 57.4; DB 4; Length 53500;

Matches 255; Conservative 0; Mismatches 291; Indels 3; Gaps 2;

QY 2329 CGAGCGGACGCTGTCTGTGTTCACGCATATGCGCCATGTCCACCGGAGCGTGGCGG 2388  
Db 20152 CGTGTGACGACAAACGAGCGGCGAGCGCGGTGATCACTATCGCGTGCAGAGCGA 20093  
QY 2389 AGAAGTGGCGCGAGCTGTGCGCTTCAGCGGAGCGCGCGCGAGGAGGCGTCTG 2448  
Db 20092 GATGACGGGCTTCTCCCGCTGCGCTGTGTGCGGAGCGCG-TGCGGGGCGGCGCGG 20034  
QY 2449 CGCTGTGCGCTGCGCGCGGCTGAGCGGAGCGCTGCGCGAGCGCTGCGCGCGCGG 2508  
Db 20033 CCGGNACCGGGGCGGCGCGGCTCTGCGCGCGGCTGTCTCTCTCGACGAGG 19974

QY 2509 AGTTCCGGGCTCTCACTACGACAGCCTCGGCAAGCTGGTCTACCTCCAGCCTGG 2568  
Db 19973 TCTGTGGGGCCCGGCTTCTCTCGCGCACACGCCGCGCGCGGAAACCGGACGCTTCA 19914  
QY 2569 TCACCGAGAGCTCCGCTGTACCCCGCGCTCTCTAGGTGAGCGCGCGACACGC--G 2626  
Db 19913 CCACGCGCGCTGCGCGAACTCCCAACCGCGCGGCTCTCTCGCGCTGTGAGAAGG 19854  
QY 2627 ACTCCGGTCCAGAGCACAGCATGAGTGGACCTGAATGCAATGCAATGCACTTG 2686  
Db 19853 CCGCCCGAGAGTCCGCGCGCGCGCGGACTCGAAACCGCTGCCCGCGCGCT 19794  
QY 2687 CCGCGCGAGACCCCAAGGGATCTCGAGAGACAGCTGTCTCGCGAGCGGACGAAGG 2746  
Db 19793 CCGAGCGCGCGCAACCGCACCGCGCGCTCTGCGCAACCGCTGAGCGCGCGTGGAGG 19734  
QY 2747 TGAGGCGCGCGGATGAGTGGAGTCTGAGTGGACCTGAATGCAATGCAATGCACTTG 2806  
Db 19733 CCATGCGCGCGCTCTCGCGAGCTGCGCGCGCTGCGCGCGCTCTGCGCGCGAAACA 19674  
QY 2807 GGGGCGCGACCGCGAGCTTCCGCGCGAGCGGTGGATCAACGAGGATGCGGCTTC 2866  
Db 19673 CCGCCCTCGACAAGCGCGATCGGAGCATACCGCGGATTCGACGAGGATTCGGGCTCA 19614  
QY 2867 GCAACGCGT 2875  
Db 19613 CTTGGGGGT 19605

## RESULT 13

US-09-192-434-1

; Sequence 1, Application US/09192434

; Patent No. 6479734

; GENERAL INFORMATION:

; APPLICANT: IBA, Koh

; APPLICANT: SHIMADA, Takiko

; APPLICANT: KUSANO, Tomonobu

; TITLE OF INVENTION: DNA FRAGMENT, RECOMBINANT DNA, AND TRANSFORMED PLANT

; FILE REFERENCE: 026350-020

; CURRENT APPLICATION NUMBER: US/09/192,434

; CURRENT FILING DATE: 1998-11-16

; PRIOR APPLICATION NUMBER: JP 10-162186

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1

; LENGTH: 3794

; TYPE: DNA

; ORGANISM: Maize breed of honey bantam

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (2798)..(3205)

US-09-192-434-1

## Query Match

Best Local Similarity 1.5%; Score 57.2; DB 4; Length 3794;

Matches 146; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

QY 2321 GCGGACACAGCGGACGACCGCTGTCTGTGTTCACGCATATGCGCCATGTCCACCGGA 2380  
Db 2869 GTGTTCACAGGAGTTCGCGCGCGGTGCGCGCGGAGAACCGCGGTGAGGA 2928  
QY 2381 CTTGCGCGAGAGCTTCGCGCGGAGCTGTGCGCTTCGAGGCGGAGCGCGCGGAGGA 2440  
Db 2929 GCTGTGCGGAGGTGCGCGCGCTTCAGCGGAGAACCGCGGTGAGGA 2988  
QY 2441 GGGCGTTCGCGCTGTGCGCTTCGCGCGGCTGACCGCGAGCAAGGCGCTTCGCGCGCG 2500  
Db 2989 GCGCGTTCGCGCGCTTCGCGCGGCTGCGCGGAGAACCGCGGTGAGGA 3048  
QY 2501 CGTGGCGAGTTCGCGCGGCTCTCTACCTACGACAGCTTCGCGCAAGCTGTCTCTCA 2560  
Db 3049 CCGCGAGCTGCGCGCGCGCTCTCTCTGCGCGCGCTCTCTCTGAGGTCTCCAGATGCG 3108

[illegible]

```

RESULT 14
US-08-682-847-1
; Sequence 1, Application US/08682847
; Patent No. 5858989
; GENERAL INFORMATION:
; APPLICANT: BABIUK, LORNE
; APPLICANT: VAN DEN HURK, SYLVIA
; APPLICANT: ZAMB, TIM
; APPLICANT: FITZPATRICK, DAVID
; TITLE OF INVENTION: RECOMBINANT BOVINE HERPESVIRUS TYPE 1
; TITLE OF INVENTION: POLYPEPTIDES AND VACCINES

```

NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 755 PAGE MILL ROAD  
CITY: PALO ALTO  
STATE: CA  
COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/682.847

FILING DATE: 12-JUL-1996  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: PARK, FREDDIE K.  
REGISTRATION NUMBER: 35,636  
REFERENCE/DOCKET NUMBER: 29310-20005.10  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 813-5600  
TELEFAX: (415) 494-0792  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3382 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)

FEATURE:  
NAME/KEY: CDS  
LOCATION: 432..3230  
IS-08-682-847-1

```

Query Match          1.4%; Score 56.4; DB 2; Length 3382;
Best Local Similarity 51.6%; Pred. No. 0.00048;
Matches 129; Conservative 0; Mismatches 121; Indels 0

```

2297	Y	GGTGTCTCAACTTCGTGTATCGCCGCGCGGGACACGACGCGCTGCTGGTTCAC	2356
1829	b	GATGCTCAGCAACGAGCTGGCCCAAGCTATCTCGAGAGCTGGCGCTCGAACGGCAC	1888
2357	Y	GCACATGGCCATGTCTCCACCCCGAGCTGGCCGAGAAAGCTGCCCGCAGCTGTGCGGGT	2416
1889	b	GCTCAGGGGGCTGTTTCGCGCGCGCGCCCCAAGCCGGGCGCGCGCGCGCGC	1948
2417	Y	CGAGGCGGAGCGCGCGCGGAGGAGGCGTCCGCTCGTGCCCTGCGCGCGCGCTGACGC	2476
1949	b	CCGCGCGCTCGCGCCCGGCGGCCGCGGCGGCGCAACGCGGCCGCGCGCGCGCAGCGC	2008
2477	Y	CGACGACAAGGCGTTTCGCGCGCCCGGTGGCGCAGTTCCGCGGGCTCTCTACCTACGACAG	2536
2009	b	CGCGCGGGGTGATACCTCGTAGCTCGCGCGAGTTTCGCGCGCTGCAGTTCACTTACGA	2068

QY	2537	CCTCGGCAAG	2546
Db	2069	CCACATCCAG	2078

```

RESULT 15
PCT-US95-04801-3
; Sequence 3, Application PC/TUS9504801
; GENERAL INFORMATION:
; APPLICANT: Martin, Juan P.
; APPLICANT: Coque, Juan R.
; APPLICANT: Esguita, Francisco J.
; APPLICANT: Fuente, Juan L.
; APPLICANT: Liarena, Francisco J.
; APPLICANT: Liras, Paloma
; TITLE OF INVENTION: DNA ENCODING CEPHAMYCIN BIOSYNTHESIS
; TITLE OF INVENTION: LATE GENES
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John W. Wallen III
; STREET: P.O. Box 2000
; CITY: Rahway
; STATE: New Jersey
; COUNTRY: USA

```

```

1  ZIP: 07065
2
3  COMPUTER READABLE FORM:
4
5  MEDIUM TYPE: Floppy disk
6
7  COMPUTER: IBM PC compatible
8
9  OPERATING SYSTEM: PC-DOS/MS-DOS
10
11  SOFTWARE: Patent in Release #1.0, Version #1.25
12
13  CURRENT APPLICATION DATA:
14
15  APPLICATION NUMBER: PCT/US95/04801
16
17  FILING DATE:
18
19  CLASSIFICATION:
20
21  ATTORNEY/AGENT INFORMATION:
22
23  NAME: Wallen III, John W.
24
25  REGISTRATION NUMBER: 35,403
26
27  REFERENCE/DOCKET NUMBER: 19179
28
29  TELECOMMUNICATION INFORMATION:
30
31  TELEPHONE: (908) 594-3905
32
33  TELEFAX: (908) 594-4720
34
35  INFORMATION FOR SEQ ID NO. 3:
36
37  SEQUENCE CHARACTERISTICS:
38
39  LENGTH: 900 base pairs
40
41  TYPE: nucleic acid
42
43  STRANDEDNESS: single
44
45  TOPOLOGY: linear
46
47  MOLECULE TYPE: cDNA
48
49  PCT-US95-04801-3

```

Query Match 1.4%; Score 56; DB 5; Length 900;  
Best Local Similarity 44.3%; Pred. No. 0.00028;

Matches	271; Conservative	0; Mismatches	340; Indels	1; Gaps	17
2244	GAGCGCGGACGACGCGGGGCTTCGGGACGACAGACCTCCGGAGCTGCTC				2303
50	GACGCGGGACGGGACTGTGATCAGACCGTGACCCGGCGCGCGAGTGCTGTC				109
2304	AATCTGTGATCCGCGCGGGACACGACGCGACGCTGTCGTGTTTCAGCACATG				2363
110	AATTTCCGCAAGTGGCGTGGAGACACCATCACCGACCTGCGGAGGCTCGATCCG				169
2364	GCCATGTCACCCGGACGTGGCCGAGAGCTGCCGCGAGCTGTGCGGTTTCGAGGG				2423
170	CCGGCGTAGACGAACAGGGGTTTGAGAAAGTCAACCGCGCCACCGCGCGTCCAGCG				229
2424	GAGCGCGCGCGAGAGGGCGTCCGCGCTGTCGCTCGCGCGCTGACGCGGACGAC				2483
230	GGCTGTGGACAGCAGAGGAAGCCCGCTGGAGCATACCGCGGGAAACCGGTGAGCTG				289
2484	AAGGGGTTCGCGCGCGTGGCGCATTCGCGGGGCTCTCACTACGACGCTCGG				2543

```
Db      290 CTCGGCTCGCTACCGCGCGGACGTGTGGAGTTCTTCGACGCCACCCTGGGGCGGACG 349
QY      2544 AAGCTGGTCTACTCTCCACGCTCGCTCACCGAGACGCTCGGCTGTACCCCGCGCTCCCT 2603
Db      350 GACGGCGCGGACGACCCCGCGCGCCAGTCCCGCACCCAGCGGTGCACGTGACACGAGC 409
QY      2604 CAGGTGAGCGCGCGGACACGCGACCTCCGGTCCAGAGCACAGCATGTGAGTGGACC 2663
Db      410 CCGGGCAGCGCGCGGCGCCAGGGCCGAGCGCACCTCGGCCCCCGCGGAGTTCGGGCGC 469
QY      2664 TGAATGCAATGCACATGCATCTTGGCGCGCGCGAGACCCCAAGGGGATCCTGGAGGACGA 2723
Db      470 TTCCAGATCATCAACGCTCTGGCGCGCGC-TGCTCGAGCGGTGCGCAACTTCCCGCTGSC 528
QY      2724 CGTGCTGCCGACGGGACGAGGTGAGGCGCGCGGATGGTGACGTAGTGGCCCTACTC 2783
Db      529 GCTGTGCGACTACCGGTGCTGGACCTGTCCGCCACCTGGTGCCGACCCCGCTGGAATT 588
QY      2784 GATGGGCGGATGGAGTACAATGGGGCCCCGACCGCGGAGCTTCCGGCGGAGCGGTG 2843
Db      589 CCCGGACTGGGTGAAGGACCGGAGAACTACTCTCGGTCCGGCACACCCCGGCGCACCGCTG 648
QY      2844 GATCAACGAGGA 2855
Db      649 GTACTTCGGGA 660
```

Search completed: March 7, 2004, 17:06:42  
Job time : 195 secs

us-10-021-657-7.rng

Mon Mar 8 10:20:24 2004

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 7, 2004, 10:14:42 ; Search time 961 Seconds  
(without alignments)  
17227.084 Million cell updates/sec

Title: US-10-021-657-7  
Perfect score: 3897  
Sequence: 1 Gaattccaagaggccctt.....gtcgagtaaacgagaattc 3897

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues  
Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N Geneseq 29Jan04 : \*  
1: Geneseqn1980s : \*  
2: Geneseqn1990s : \*  
3: Geneseqn2000s : \*  
4: Geneseqn2001as : \*  
5: Geneseqn2001bs : \*  
6: Geneseqn2002as : \*  
7: Geneseqn2003as : \*  
8: Geneseqn2003bs : \*  
9: Geneseqn2003cs : \*  
10: Geneseqn2004s : \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3897	100.0	3897	6	ABL60615 Maize SBM
2	1378.4	35.4	1906	6	ABL60611 Maize SBM
3	1092	28.0	1092	6	ABL60612 Sorghum S
4	268.4	6.9	494	6	ABL60614 Maize SBM
5	263	6.7	267	6	ADA70596 Rice gene
6	137.8	3.5	1527	7	ADA71342 Rice gene
7	127	3.3	1536	6	ACL21310 DNA clone
8	120.8	3.1	657	8	ACL21310 Codon mod
9	119.6	3.1	1626	2	AAX26878 Cytochrom
10	119.6	3.1	2181	2	AAX26874 Rice gene
11	114	2.9	1563	7	ACL21306 DNA clone
12	112.8	2.9	621	8	ADA49310 Maize gene
13	105.4	2.7	528	8	ACL21305 DNA clone
14	104	2.7	399	8	ACL21309 DNA clone
15	104	2.7	460	8	ACL21308 DNA clone
16	104	2.7	507	8	ACL21304 DNA clone
17	104	2.7	527	8	ABZ13076 Arabidops
18	104	2.7	1662	6	ADA71125 Rice gene
19	100.6	2.6	1965	7	ADA70631 Rice gene
20	99.4	2.6	1506	7	ABZ14873 Arabidops
21	96	2.5	1614	6	AAC45033 Arabidops
22	96	2.5	1896	3	AAC45033 Arabidops
23	93.8	2.4	866	6	ABN98804 Arabidops

24	93.2	2.4	1395	8	ADA48303 Rice gene
25	92.4	2.4	1542	6	ABZ14063 Arabidops
26	92.4	2.4	1736	2	AAX58407 DNA seque
27	90	2.3	1473	6	ABZ14176 Arabidops
28	86	2.2	843	8	ADA48919 Wheat gen
29	81.6	2.1	1694	3	AAC37317 Arabidops
30	81.4	2.1	1488	6	ABZ13809 Arabidops
31	81.4	2.1	1488	7	ADA68342 Arabidops
32	79	2.0	1440	3	AAC42927 Arabidops
33	78.6	2.0	589	3	AAC38695 Arabidops
34	78.4	2.0	1117	3	AAC40695 Arabidops
35	78.4	2.0	1533	3	AAC42704 Banana ge
36	78.2	2.0	1537	8	ADA48779 Arabidops
37	75.2	1.9	1551	3	AAC43122 Arabidops
38	74.4	1.9	665	6	ABQ65871 Arabidops
39	74	1.9	1407	3	AAC33115 Arabidops
40	72.2	1.9	678	8	ACL21311 DNA clone
41	70.8	1.8	31263	7	ACA37577 Prokaryot
42	68.6	1.8	818	4	AAH06808 Human CDN
43	68.6	1.8	1961	4	AAX14525 Human sec
44	68.2	1.8	3871	2	AAX60801 Human sec
45	68.2	1.8	3871	4	AAS9207 Human.CDN

ALIGNMENTS

RESULT 1  
ABL60615 standard; DNA; 3897 BP.  
ID ABL60615 standard; DNA; 3897 BP.  
XX  
AC ABL60615;  
XX  
DT 27-AUG-2002 (first entry)  
XX  
DE Maize SBMu200 genomic DNA sequence.  
XX  
KW SBMu200; fertility; maize; hybrid seed; sterility; plant; gene;  
XX  
KW transgenic; ds.  
XX  
OS Zea mays.  
XX  
PN WC200226789-A2.  
XX  
PD 04-APR-2002.  
XX  
PF 25-SEP-2001; 2001WO-US029886.  
XX  
PR 26-SEP-2000; 2000US-00670153.  
XX  
XX (PION-) PIONEER HI-BRED INT INC.  
XX  
XX Albertsen MC, Fox T, Huffman G, Trimmell M;  
XX  
XX WPI; 2002-471258/50.  
XX  
XX Novel nucleic acid sequence comprising SBMu200 gene, useful for producing  
XX transgenic plants and mediating male fertility in plants.  
XX  
XX Claim 3; Fig 5; 60pp; English.  
XX  
XX The invention relates to a SBMu200 gene, useful for impacting male  
XX fertility in a plant e.g. maize. An isolated nucleotide sequence  
XX comprising the SBMu200 gene is useful for producing hybrid seeds and for  
XX providing heritable male sterility in a plant.  
XX The present sequence represents the maize SBMu200 genomic DNA sequence  
XX  
XX Sequence 3897 BP; 927 A; 1013 C; 1058 G; 899 T; 0 U; 0 Other;  
XX  
XX Query Match 100.0%; Score 3897; DB 6; Length 3897;  
XX Best Local Similarity 100.0%; Pred. No. 0;  
XX Matches 3897; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

[illegible]





```

QY 2768 CGTACGTGCTTACTCGATGGCGGATGAGTACAACTGGGCCCCGACGCGCGAGCT 2827
Db 1328 CGTACGTGCTTACTCGATGGCGGATGAGTACAACTGGGCCCCGACGCGCGAGCT 1387
QY 2828 TCCGGCCGAGCGGTGATCAACGAGATGGCGGCTTCCGCAACGGCTGCGGTTCAAGT 2887
Db 1388 TCCGGCCGAGCGGTGATCAACGAGATGGCGGCTTCCGCAACGGCTGCGGTTCAAGT 1447
QY 2888 TCACGGCGTTCACGCGCGGCGAGGATCTGCTGGGCAAGGACTCGGCGTACTGCGAGA 2947
Db 1448 TCACGGCGTTCACGCGCGGCGAGGATCTGCTGGGCAAGGACTCGGCGTACTGCGAGA 1507
QY 2948 TGAAGATGGCGCTGGCCATCCTCTTGGCTTCTACAGCTTCCGGCTGCTGGAGGGGACC 3007
Db 1508 TGAAGATGGCGCTGGCCATCCTCTTGGCTTCTACAGCTTCCGGCTGCTGGAGGGGACC 1567
QY 3008 CGGTGCAGTACCGCATGATGACCATCTCTCCATGGCGGCGGCTCAAGTCCGGCTCT 3067
Db 1568 CGGTGCAGTACCGCATGATGACCATCTCTCCATGGCGGCGGCTCAAGTCCGGCTCT 1627
QY 3068 CTAGGGCGCTGATGTCATGGCGATTG-----GGATATCATCCCGCTTAATCC----- 3117
Db 1628 CTAGGGCGCTGATGTCATGGCGATTG-----GGATATGATATCGTCCCGCTTAATCCAGAC 1687
QY 3118 -----TTAAATTTGCGATGCGATGTAAGGAAAGCGATGGGTTTCAT 3163
Db 1688 AATAACGCTCGTGTATCAAAATTTGCGATGCGATGTAAGGAAAGCGATGGGTTTCAT 1747
QY 3164 TGGTGGCTTGGCTTAAGCCTTAAACCTCGCTGGGCTTTCGGAACACACATCAGTAG 3223
Db 1748 TGGTGGCTTGGCTTAAGCCTTAAACCTCGCTGGGCTTTCGGAACACACATCAGTAG 1807
QY 3224 TGTTTTGACTCTACTCTCAGTGAAGTGTAGTGACAGCATACAAGTTTCATCATATA 3283
Db 1808 TGTTTTGACTCTACTCTCAGTGAAGTGTAGTGACAGCATACAAGTTTCATCATATA 1867
QY 3284 TTATCCTCTTTCTT 3297
Db 1868 TTATCCTCTTTCTT 1881

RESULT 3
ABL60613
ID ABL60613 standard; DNA; 1092 BP.
AC ABL60613;
XX 27-AUG-2002 (first entry)
DE Maize SBMu200 promoter sequence.
XX SBMu200; gene; fertility; maize; hybrid seed; sterility; plant; promoter;
KW transgenic; ds.
XX Zea mays.
OS WO200226789-A2.
XX PN 04-APR-2002.
XX PF 25-SEP-2001; 2001WO-US029986.
XX PR 26-SEP-2000; 2000US-00670153.
XX PA (PION-) PIONEER HI-BRED INT INC.
XX PI Albertsen MC, Fox T, Huffman G, Trimmell M;
XX DR WPI; 2002-471258/50.
XX PT Novel nucleic acid sequence comprising SBMu200 gene, useful for producing
transgenic plants and mediating male fertility in plants.

```

```

XX Claim 36; Fig 8; 60pp; English.
XX The invention relates to a SBMu200 gene, useful for impacting male
fertility in a plant e.g. maize. An isolated nucleotide sequence
comprising the SBMu200 gene is useful for producing hybrid seeds and for
providing heritable externally controllable male sterility in a plant.
XX The present sequence represents the maize SBMu200 promoter sequence
SQ Sequence 1092 BP; 313 A; 208 C; 236 G; 335 T; 0 J; 0 Other;

Query Match 28.0%; Score 1092; DB 6; Length 1092;
Best Local Similarity 100.0%; Pred. No. 6.7e-234;
Matches 1092; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCGAAGCGAGGCGCTTGTAGCAGAGAGTGTGTGTGATGCGGCGGAAATGAGT 60
Db 1 GAATTCGAAGCGAGGCGCTTGTAGCAGAGAGTGTGTGTGATGCGGCGGAAATGAGT 60
QY 61 GCGTGTGAGAGCAACGCTGAGGGGTTCCAGGATGCAATGGCTATGGCAATCGGCTAG 120
Db 61 GCGTGTGAGAGCAACGCTGAGGGGTTCCAGGATGCAATGGCTATGGCAATCGGCTAG 120
QY 121 AGGTGGAGGACAAAGGTGGTGGAGGATTTGGAGGGGCACTATGGCAAGTTGGTGAAGAGGC 180
Db 121 AGGTGGAGGACAAAGGTGGTGGAGGATTTGGAGGGGCACTATGGCAAGTTGGTGAAGAGGC 180
QY 181 ACGCAATGAGAGATCTATTTCAGACTTACACTGGATGCGGCAACAAATTCACCTTTAGA 240
Db 181 ACGCAATGAGAGATCTATTTCAGACTTACACTGGATGCGGCAACAAATTCACCTTTAGA 240
QY 241 TTTTGATCTGTCACTCTACTTATTTCTGTTGGTGGCAACTTCCAAATAGGCTCATGTT 300
Db 241 TTTTGATCTGTCACTCTACTTATTTCTGTTGGTGGCAACTTCCAAATAGGCTCATGTT 300
QY 301 AATCAATGATTAGTGATTATTTCAGCAAAATATTCTTGTGTTGACATTTATATATGTG 360
Db 301 AATCAATGATTAGTGATTATTTCAGCAAAATATTCTTGTGTTGACATTTATATATGTG 360
QY 361 GGGTGAGACGGATTAATAATATCATCCATGAGAGCTTTATCTTCATGCTCTCTGATTTGG 420
Db 361 GGGTGAGACGGATTAATAATATCATCCATGAGAGCTTTATCTTCATGCTCTCTGATTTGG 420
QY 421 TTTTCAGATCATTCTTTTCAGTGTTCACAGAAATTTTCTCAGTTTGGTCCATGATTTTGG 480
Db 421 TTTTCAGATCATTCTTTTCAGTGTTCACAGAAATTTTCTCAGTTTGGTCCATGATTTTGG 480
QY 481 AAGTGAGGTTCTTTAAATTTTCAATTTATGCTTCTCTTTCTTTCTAGACTAGCACTGATGA 540
Db 481 AAGTGAGGTTCTTTAAATTTTCAATTTATGCTTCTCTTTCTTTCTAGACTAGCACTGATGA 540
QY 541 CTTTTCACCTTTGGGTTTCAAAATTTGACTCACAAGAAAACAAATTCACCTTTGGGTTTACA 600
Db 541 CTTTTCACCTTTGGGTTTCAAAATTTGACTCACAAGAAAACAAATTCACCTTTGGGTTTACA 600
QY 601 AATTCTCTTTCAGGATGTTTTCAGTGAAGTGTCTATGATAGGAAACAGGAAATGGCT 660
Db 601 AATTCTCTTTCAGGATGTTTTCAGTGAAGTGTCTATGATAGGAAACAGGAAATGGCT 660
QY 661 CAGTTTTTTAAGGAACAATGTACAGATTTTCAATTTTCAGAACTCTTTCTGGTTGGTGGTTT 720
Db 661 CAGTTTTTTAAGGAACAATGTACAGATTTTCAATTTTCAGAACTCTTTCTGGTTGGTGGTTT 720
QY 721 CAGACTTTTTGTACCAAGCTGATGGATCAAAATTTGTTTCCAAAGTCTGATTAACAGAA 780
Db 721 CAGACTTTTTGTACCAAGCTGATGGATCAAAATTTGTTTCCAAAGTCTGATTAACAGAA 780
QY 781 ACTGGCAACTCCTAATTGATATAATAAAGATATAAATACAGTATCAGATATCTCATTTTC 840
Db 781 ACTGGCAACTCCTAATTGATATAATAAAGATATAAATACAGTATCAGATATCTCATTTTC 840
QY 841 TTGGTTGGCAGATCACAAGAAAGGAAACAAAGGCTTAAGCCCTCTTCTGTTGGGAGTTA 900
Db 841 TTGGTTGGCAGATCACAAGAAAGGAAACAAAGGCTTAAGCCCTCTTCTGTTGGGAGTTA 900

```



CC fertility in a plant e.g. maize. An isolated nucleotide sequence  
CC comprising the SBM200 gene is useful for producing hybrid seeds and for  
CC providing heritable externally controllable male sterility in a plant.  
CC The present sequence represents an essential region of the maize SBM200  
CC promoter sequence

XX Sequence 267 BP; 81 A; 66 C; 54 G; 66 T; 0 U; 0 Other;

Query Match 6.7%; Score 263; DB 6; Length 267;  
Best Local Similarity 100.0%; Pred. No. 8.9e-49;  
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 830 AACTCATTTCTTGTGTCAGATCATCAAAAGGACACAAAGGCTAGCCCTCCTACTTG 889  
DB 5 ATCTCATTTTCTTGTGTCAGATCATCAAAAGGACACAAAGGCTAGCCCTCCTACTTG 64  
QY 890 TTCCGGAGTTAGTCAAGGACACCATATCAATGAAAGAAATCTTAATTGGGGTCAAC 949  
DB 65 TTCCGGAGTTAGTCAAGGACACCATATGATGAAGAAATCTTAATTGGGGTCAAC 124  
QY 950 AAGATTGTCTCTCTCGAGTTGGGGGTCCTTAAGGTTGGTAGTACCAATACCCAAATATA 1009  
DB 125 AAGATTGTCTCTCTCGAGTTGGGGGTCCTTAAGGTTGGTAGTACCAATACCCAAATATA 184  
QY 1010 TCACCTAACAAACCCATCCATGCTACATACATACATACATACATCCATCCTTGTAGACTGG 1069  
DB 185 TCACCTAACAAACCCATCCATGCTACATACATACATACATACATCCATCCTTGTAGACTGG 244  
QY 1070 ACCCTTCATCAAGAGCACCATGG 1092  
DB 245 ACCCTTCATCAAGAGCACCATGG 267

## RESULT 6

ADA70596  
ID ADA70596 standard; DNA; 1527 BP.

XX AC ADA70596;

XX 20-NOV-2003 (first entry)

XX Rice gene, SEQ ID 3919.

XX Plant; bacterial infection; fungal infection; viral infection; rice;  
KW Gene; ds.

XX Oryza sativa.

XX WO2003000898-A1.

XX 03-JAN-2003.

XX 22-JUN-2001; 2001WO-IB001105.

XX 22-JUN-2001; 2001WO-IB001105.

XX (SYGN ) SYNGENTA PARTICIPATIONS AG.

XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;  
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;

XX WPI; 2003-175290/17.

XX Identifying at least one gene involved in plant resistance or response to  
PT pathogenic infection for conferring resistance or tolerance to a plant to  
PT bacterial, fungal or viral infection by determining or detecting plant  
PT gene expression.

XX Claim 6; SEQ ID NO 3919; 899pp; English.

XX The present invention relates to a method (M1) for identifying genes  
CC involved in plant resistance or response to pathogenic infection. M1  
CC comprises identifying a gene whose expression is significantly altered in

CC the incompatible interaction of plant gene expression relative to  
CC expression of the gene in an uninfected plant, in a mutant plant that  
CC does not express a gene associated with response to pathogenic infection,  
CC or in a corresponding incompatible or compatible interaction. (M1) is  
CC useful for conferring resistance to resistance or tolerance to a plant to  
CC bacterial, fungal or viral infection. The present sequence was used to  
CC illustrate the invention.

XX Sequence 1527 BP; 280 A; 473 C; 494 G; 280 T; 0 U; 0 Other;

Query Match 3.5%; Score 137.8; DB 7; Length 1527;  
Best Local Similarity 70.5%; Pred. No. 1.7e-20;  
Matches 184; Conservative 0; Mismatches 77; Indels 0; Gaps 0;  
QY 2692 GCGCAGGACCCCAAGGGGATCTTGAGGACGACGCTGCTGCGGACGGGACGAGGTGAGG 2751  
DB 1135 GCGATGAGTCTGCGACAGCTGCAAGGAGGACGACTTCTTCCGCGGACGGCACGTTCTGGGG 1194  
QY 2752 GCGCGGGGATGGTACGCTAGTGCCTTACTCGATGGGGCGGATGAGTACACTGGGGC 2811  
DB 1195 AAAGGTGGTTGTGTCTCTACAGCGCTACGCGATGGCAGGATGGAGGACATCTGGGGC 1254  
QY 2812 CCGCAGCGGGGAGCTTCCGGCGGAGCGGTGGATCAACGAGGATGGCGCGTCCGCAAC 2871  
DB 1255 GGGACTGTGAGGAGTTTCAAGGACAGCGGTGGCTAGACGAGCGCGCGTTCGGCGG 1314  
QY 2872 GCGTCGCGGTTCAAGTTCAAGGCGTTCCAGGCGGGCGGAGGATCTGCTGGGCAAGGAC 2931  
DB 1315 GAGAGCGCGTTCAAGTACCGGTTTCCACGCGGGCGGAGGATGTGCTCGGCAAGGAG 1374  
QY 2932 TCGGCGTACCTGCCATGAAG 2952  
DB 1375 ATGGCTACATACATGAAG 1395

## RESULT 7

ADA71342  
ID ADA71342 standard; DNA; 1536 BP.

XX AC ADA71342;

XX 20-NOV-2003 (first entry)

XX Rice gene, SEQ ID 4665.

XX Plant; bacterial infection; fungal infection; viral infection; rice;  
KW Gene; ds.

XX Oryza sativa.

XX WO2003000898-A1.

XX 03-JAN-2003.

XX 22-JUN-2001; 2001WO-IB001105.

XX 22-JUN-2001; 2001WO-IB001105.

XX (SYGN ) SYNGENTA PARTICIPATIONS AG.

XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;  
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;

XX WPI; 2003-175290/17.

XX Identifying at least one gene involved in plant resistance or response to  
PT pathogenic infection for conferring resistance or tolerance to a plant to  
PT bacterial, fungal or viral infection by determining or detecting plant  
PT gene expression.

XX Claim 6; SEQ ID NO 4665; 899pp; English.

XX The present invention relates to a method (M1) for identifying genes

CC involved in plant resistance or response to pathogenic infection. M1  
CC comprises identifying a gene whose expression is significantly altered in  
CC the incompatible interaction of plant gene expression relative to  
CC expression of the gene in an uninfected plant, in a mutant plant that  
CC does not express a gene associated with response to pathogenic infection,  
CC or in a corresponding incompatible or compatible interaction. (M1) is  
CC useful for conferring resistance to resistance or tolerance to a plant to  
CC bacterial, fungal or viral infection. The present sequence was used to  
CC illustrate the invention.

XX  
SQ Sequence 1536 BP; 301 A; 478 C; 479 G; 278 T; 0 U; 0 Other;

Query Match 3.3%; Score 127; DB 7; Length 1536;  
Best Local Similarity 68.6%; Pred. No. 4.3e-18;  
Matches 175; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

Qy 2698 GACCCCAAGGGATCTTGGAGGACGCTGCTGCGGACGGGACGAGGAGGCGCGGC 2757

Db 1147 GACACGCACAGCTGACAGGAGGACGCTTCTGCGGACGGCAGCAGTTCGTTGGGGAAGGG 1206

Qy 2758 GCGATGTTGACCTACGTCCTTCTGCGGACGGGATGAGTACAACTGGGGGCGCGAC 2817

Db 1207 TGGCTAGTGTACAGCGCGTATGCCATGGGCGTATGAGGCGATCTGGGGCGGAC 1266

Qy 2818 GCGGCGAGCTTCGCGCGGACGGTGTGATCAACGAGGATGCGCGTTCCGCAACGCGTCG 2877

Db 1267 TCGGAGGAGTACAGCGCGGAACGGTGGCTGGACGAGGCGGCGGCTTCGCGCGGAGAGC 1326

Qy 2878 CCGTTCAAGTTACGGCGCTTCCAGCGGGCGGAGGATCTGCTGGGCAAGGACTCGGCG 2937

Db 1327 ACGTTCAAGTACCGCGTGTTCACGCGCGGCTTGGATCTGATCTGCGGACGAGATGCGC 1386

Qy 2938 TACCTGCAGATGAAG 2952  
Db 1387 TACATACAATGAAG 1401

RESULT 8  
ACL21310/C  
ID ACL21310 standard; DNA; 657 BP.

XX  
AC ACL21310;  
XX  
XX  
DT 27-OCT-2003 (revised)  
DT 17-OCT-2003 (first entry)

XX  
DE DNA clone originating in barley containing SNP encoding sequence #11301.

XX  
KW Barley; single nucleotide polymorphism; SNP; genotype-phenotype analysis;  
KW gene; ss.

XX  
OS Hordeum vulgare; var. (cul.Akashinriki).

XX  
PN WO2003057877-A1.

XX  
PD 17-JUL-2003.

XX  
PF 16-DEC-2002; 2002WO-IB005403.

XX  
XX 20-DEC-2001; 2001JP-00387059.

XX  
PR 20-DEC-2001; 2001JP-00387131.

XX  
PR 20-DEC-2001; 2001JP-00403299.

XX  
PR 20-DEC-2001; 2001JP-00403300.

XX  
PR 27-SEP-2002; 2002JP-00327515.

XX  
PA (UYN1-) UNIV JAPAN OKAYAMA.

XX  
PI Sato K, Takeda K, Kohara Y;

XX  
DR WPI; 2003-587127/55.

XX  
PT Single nucleotide polymorphism sites in barley varieties and DNA  
PT sequences containing them for analysis and identification of barley

PT varieties and production of barley transformants with desired  
PT characteristics.

XX  
XX  
XX Disclosure; SEQ ID XX; 284pp; Japanese.

XX  
CC The present invention relates to oligonucleotide clones originating in  
CC barley (Hordeum vulgare) which contain single nucleotide polymorphisms  
CC (SNP). The oligonucleotides may be used for analysis of SNPs among barley  
CC varieties, identification of particular varieties and genotype-phenotype  
CC analysis, isolation of specific genes and creation of new varieties by  
CC transformation of barley varieties with them and production of new barley  
CC varieties with desired properties. The present sequence represents an  
CC oligonucleotide clone DNA sequence featured in the specification. The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published-pct-sequences. (Updated on 27-OCT-2003 to  
CC standardise OS field)

XX  
SQ Sequence 657 BP; 140 A; 227 C; 159 G; 131 T; 0 U; 0 Other;

Query Match 3.1%; Score 120.8; DB 8; Length 657;  
Best Local Similarity 60.7%; Pred. No. 7.7e-17;  
Matches 216; Conservative 0; Mismatches 137; Indels 3; Gaps 1;

Qy 2713 CTGAGGACACGCTGCTGCGGACGGACGAAAGTGAGGCGCGCGGATGTCGACGTAC 2772

Db 560 CAGGCCCGACGCTTCTTCGGACGGCACCGCGGTGGGAGCGGTTGGTTCTGGCGGTAC 501

Qy 2773 GTGCGCTTCTGATGGGCGGATGGAGTCAACTGGGGCGCGGACGCTTCGCGG 2832

Db 500 AACTGCTATGCATGGGCGGATGGAGTCTGTGTGGGCGGAGCGCGCAGGCGTACCGG 441

Qy 2833 CCGGAGCGGTGGATCAAC---GAGGATGGCGGCTTCGCAACGCGTCCGCGTTCAGTTC 2889

Db 440 CCGGAGCGGTGGTGGACCGCGGCGGAGGACGTTCCGCGCGGAGAGCCGTTCCGCGTAC 381

Qy 2890 ACGCGTTCCAGCGGGCGGAGGATCTGCTGGCAAGGACTCGGCGTACCTGCAGATG 2949

Db 380 ATGGCGTTTCACGCGGGCGCAAGAAATTGCTTCGAAAGGAGATGGCGTATATCCAGATG 321

Qy 2950 AAGATGGCGTGGCCATCTCTTCGCGCTTCTACAGCTTCGCGTTCGAGGGGACCCCG 3009

Db 320 AAGTCTATCGTGGCGTGGTGGAGAGTTCGAAATGGCGGTGGACGGCGGTACCGG 261

Qy 3010 GTGCAGTACCGCATGATGACCATCTCTCCATGGCGCACGCGCTCAAGGTCGCGGT 3065

Db 260 CCGCGGAGGTGGCATCACTCACTTACGAATGGCGGACGGGCTCCCGGTGCGGGT 205

RESULT 9  
AA26878  
ID AAX26878 standard; cDNA; 1626 BP.

XX  
XX  
XX AAX26878;

XX  
XX  
XX 23-JUN-1999 (first entry)

XX  
XX  
XX Codon modified cDNA encoding a cytochrome P450.

XX  
XX  
XX Cytochrome P450; protein expression; yeast; mutant; ss.

XX  
XX  
XX Synthetic.

XX  
XX  
XX Triticum aestivum.

XX  
XX  
XX Key Location/Qualifiers

XX  
XX  
XX CDS 1..1626

XX  
XX  
XX FT /\*tag= a

XX  
XX  
XX FR2768748-A1.

XX  
XX  
XX 26-MAR-1999.

XX  
XX  
XX 24-SEP-1997; 97FR-00012094.

```
XX PR 24-SEP-1997; 97FR-00012094.
XX PA (RHON ) RHONE-POULENC AGROCHIMIE.
XX PI Batard Y, Schalk M, Durst F, Werck RD;
XX WPI: 1999-217499/19.
XX P-PSDB; AAY01648.
XX DNA for expression in yeasts - with codon changes based on yeast codon
XX usage.
XX Example 4; Page 24-26; 31pp; French.
XX CC The present sequence encodes a modified cytochrome P450, which was
XX created by altering the codons of AAX26874 to exemplify the invention.
XX CC The specification describes a DNA sequence that encodes a protein of
XX interest and contains regions with a high content of codons poorly suited
XX to yeasts, where a sufficient number of these codons are replaced by
XX corresponding codons suited to yeasts in these regions. Yeasts
XX transformed with vectors such DNA sequences can be cultured to produce
XX the protein of interest (especially a plant protein) or, when the protein
XX is an enzyme (such as cytochrome P450), can be cultured in the presence
XX of a substrate for the enzyme to produce a substrate conversion product
XX
XX SQ Sequence 1626 BP; 250 A; 583 C; 514 G; 279 T; 0 U; 0 Other;
Query Match 3.1%; Score 119.6; DB 2; Length 1626;
Best Local Similarity 59.8%; Pred. No. 2e-16;
Matches 219; Conservative 0; Mismatches 144; Indels 3; Gaps 1;
QY 2691 CGCGGAGGAGCCCAAGGGATCTCTGGAGGACGACGTGCTCCGGAGCGGACGAGGTGAG 2750
DB 1149 CCCCGAGGACTCCAAAGCAGCTGCTCGCGGACGACTACCTCCCGGACGACCTTCGTGCC 1208
QY 2751 GGCGCGCGGGATGGTGACGTACGTGCTTACTCGATGGGCGGATGGAGTACAACTGGGG 2810
DB 1209 GGCGGGGTGCTGGTCACTTACTTCCATATCTCGCGGGGCGCATGAAGGGGTGTGGGG 1268
QY 2811 CCCCGAGCGGCGAGCTTCGCGCGGAGCGGTGGATCAACGAGGATGGGCGC---GTTCCG 2867
DB 1269 GGAGGACTGCTCGAGTTCGCGCGGAGCGATGGCTGTCGCGGACGGCACCAAGTTCGA 1328
QY 2868 CAACGGCTGCCGCTTCAAGTTCAGCGGTTCCAGGCGGGGCGGAGGATTCGCTGGGCAA 2927
DB 1329 GCAGCAGCACTCGTCAAGTTCGTGGCGTTCAACGCGCGGGCGGAGGTGTGCTGGGCAA 1388
QY 2928 GGACTCGGGTACCTGCAGATGAAGATGGGCTGGGCGCATCTCTTGGCTTCTACAGCTT 2987
DB 1389 GGAACCTAGCTTACCTGCAGATGAAGAACATCGCGGGGAGGTGCTGCTCGGCACCGCCT 1448
QY 2988 CCGGCTGTGGAGGGGACCCCGGTGCAGTACCGCATGATGACCATCTCTTCCATGGCGCA 3047
DB 1449 GACCGTGGCGCGGCGCCACCGCGTGGAGCAGAAGATGTGCTCAGCTCTTTCATGAAGG 1508
QY 3048 CGGCTT 3053
DB 1509 CGGGCT 1514
```

## RESULT 10

```
ANX26874
ID AAX26874 standard; cDNA; 2181 BP.
XX AC
XX AAX26874;
XX DT
XX 23-JUN-1999 (first entry)
XX CYtochrome P450 cDNA designated CYP8A5.
XX KW Cytochrome P450; protein expression; yeast; ss.
XX
```

```
OS Triticum aestivum.
XX FH Key Location/Qualifiers
XX CDS 112..1737
XX FT /*tag= a
XX PN FR2768748-A1.
XX PD 26-MAR-1999.
XX PF 24-SEP-1997; 97FR-00012094.
XX PR 24-SEP-1997; 97FR-00012094.
XX PA (RHON ) RHONE-POULENC AGROCHIMIE.
XX PI Batard Y, Schalk M, Durst F, Werck RD;
XX WPI: 1999-217499/19.
XX P-PSDB; AAY01647.
XX DNA for expression in yeasts - with codon changes based on yeast codon
XX usage.
XX Claim 20; Page 21-23; 31pp; French.
XX CC The present sequence encodes a cytochrome P450, and was used to exemplify
XX the invention. The specification describes a DNA sequence that encodes a
XX protein of interest and contains regions with a high content of codons
XX poorly suited to yeasts, where a sufficient number of these codons are
XX replaced by corresponding codons suited to yeasts in these regions.
XX CC Yeasts transformed with vectors such DNA sequences can be cultured to
XX produce the protein of interest (especially a plant protein) or, when the
XX protein is an enzyme (such as cytochrome P450), can be cultured in the
XX presence of a substrate for the enzyme to produce a substrate conversion
XX product
XX
XX SQ Sequence 2181 BP; 395 A; 722 C; 684 G; 380 T; 0 U; 0 Other;
Query Match 3.1%; Score 119.6; DB 2; Length 2181;
Best Local Similarity 59.8%; Pred. No. 2.2e-16;
Matches 219; Conservative 0; Mismatches 144; Indels 3; Gaps 1;
QY 2691 CGCGGAGGAGCCCAAGGGATCTCTGGAGGACGACGTGCTCCGGAGCGGACGAGGTGAG 2750
DB 1260 CCCCGAGGACTCCAAAGCAGCTGCTCGCGGACGACTACCTCCCGGACGACCTTCGTGCC 1319
QY 2751 GGCGCGCGGGATGGTGACGTACGTGCTTACTCGATGGGCGGATGGAGTACAACTGGGG 2810
DB 1320 GGCGGGGTGCTGGTCACTTACTTCCATATCTCGCGGGGCGCATGAAGGGGTGTGGGG 1379
QY 2811 CCCCGAGCGGCGAGCTTCGCGCGGAGCGGTGGATCAACGAGGATGGGCGC---GTTCCG 2867
DB 1380 GGAGGACTGCTCGAGTTCGCGCGGAGCGATGGCTGTCGCGCGGACGGCACCAAGTTCGA 1439
QY 2868 CAACGGCTGCCGCTTCAAGTTCACGGCTTCAGGCGGGGCGGAGGATTCGCTGGGCAA 2927
DB 1440 GCAGCAGCACTCGTACAAAGTTCGTGGCGTTCAACGCGCGGGCGGAGGTGTGCTGGGCAA 1499
QY 2928 GGACTCGGGGTACCTGCAGATGAAGATGGCGCATCTCTTGGCTTCTTACAGCTT 2987
DB 1500 GGAACCTAGCTTACCTGCAGATGAAGAACATCGCGGGGAGGTGCTGCTCGGCACCGCCT 1559
QY 2988 CCGGCTGTGGAGGGGACCCCGGTGCAGTACCGCATGATGACCATCTCTTCCATGGCGCA 3047
DB 1560 GACCGTGGCGCGGCGCCACCGCGTGGAGCAGAAGATGTGCTCAGCTCTTTCATGAAGG 1619
QY 3048 CGGCTT 3053
DB 1620 CGGGCT 1625
```

## RESULT 11

ADA70663  
 ID ADA70663 standard; DNA; 1563 BP.  
 XX AC  
 XX AC ADA70663;  
 XX XX  
 DT 20-NOV-2003 (first entry)  
 XX XX  
 DE Rice gene, SEQ ID 3986.  
 XX XX  
 KW plant; bacterial infection; fungal infection; viral infection; rice;  
 KW gene; ds.  
 XX XX  
 OS Oryza sativa.  
 XX XX  
 FN WO2003000898-A1.  
 XX XX  
 PD 03-JAN-2003.  
 XX XX  
 EF 22-JUN-2001; 2001WO-IB001105.  
 XX XX  
 PR 22-JUN-2001; 2001WO-IB001105.  
 XX XX  
 PA (SYGN) SYNGENTA PARTICIPATIONS AG.  
 XX XX  
 PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;  
 PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;  
 XX XX  
 DR WPI; 2003-175290/17.  
 XX XX  
 PT Identifying at least one gene involved in plant resistance or response to  
 PT pathogenic infection for conferring resistance or tolerance to a plant to  
 PT bacterial, fungal or viral infection by determining or detecting plant  
 PT gene expression.  
 XX XX  
 PS Claim 6; SEQ ID NO 3986; 899pp; English.  
 XX XX  
 CC The present invention relates to a method (M1) for identifying genes  
 CC involved in plant resistance or response to pathogenic infection. M1  
 CC comprises identifying a gene whose expression is significantly altered in  
 CC the incompatible interaction of plant gene expression relative to  
 CC expression of the gene in an uninfected plant, in a mutant plant that  
 CC does not express a gene associated with response to pathogenic infection,  
 CC or in a corresponding incompatible or compatible interaction. (M1) is  
 CC useful for conferring resistance to resistance or tolerance to a plant to  
 CC bacterial, fungal or viral infection. The present sequence was used to  
 CC illustrate the invention.  
 XX XX  
 SQ Sequence 1563 BP; 221 A; 572 C; 530 G; 240 T; 0 U; 0 Other;  
 Query Match 2.9%; Score 114; DB 7; Length 1563;  
 Best Local Similarity 52.6%; Fred. No. 3.5e-15;  
 Matches 320; Conservative 0; Mismatches 280; Indels 8; Gaps 3;  
 QY 2406 CTGTGCGGCTTCGAGCGGAGCGCGCGCGGAGGAGGCGTGTGCTGCTGCGCGC 2465  
 DB 850 CTGTGCGGCTTCATGGCTTCCACCGGCTGCGAGCGCGCGCGGAGGACAGTTCCTC 909  
 QY 2466 GCGCTGACGCGGACAGAGCGGTTCGCGCGCGCGGTGCGCGAGTTCGCGGCTTCCTC 2525  
 DB 910 CGCGACATGCTGTCAGCTTCTCTCTGCGCGCGGCGGACACGCTGTCCACCGCTCACC 969  
 QY 2526 ACCTAGCAGCTCGGCAAGTGTCTTACCTCCAGCTGTGTACCGAGAGCTCCGC 2585  
 DB 970 AGCTGTTTCATGCTCTCTGTCCAGAACCCCGAGTGTGCGCGCGCATGCGCGGAGGCG 1029  
 QY 2586 CTGTACCCCGCGCTCTCCTCAGGTGAGCGCGCGCGCGACCGACCTTCGCTCCAGAG---- 2641  
 DB 1030 GAGCGCGGAGCGGCGGAGAGCGCGCGCGCGCATCCTAGAGACCTTAGGGCCTC 1089  
 QY 2642 CACAGCATGCAATGAGTGAAGTGAATGCAATGCAATGCAATGCAATGCAATGCAATG 2700  
 DB 1090 CACTACACCCAGCGCGTGTGTCACGAGAACATGCGGCTGTTCGCCGCGGTGCAATTCGAC 1149

2701 CCCAAGGGGATCCTGGAGGACGACGCTGCTCCCGGACCGGACGGAAGTGAGGGCCGCGCGG 2760  
 DB 1150 TCCAAGTTTTCGCGCGCGCGGACGCTGCTCCCGGACCGGACCTACGTCGCGCGGACGCG 1209  
 QY 2761 ATGCTGACGTACGTGCTGCTGATGGGCGGATGAGTACAACTGGGCGCCCGGACGCG 2820  
 DB 1210 CGGCTATGTACACCCCTACGCAATGGGCGGATGCCGACATCTGGGCGCGGATAC 1269  
 QY 2821 GCGAGCTTCGCGGCGGAGCGGTGATCAACGAGGATGGCG---CGTTCCGCAAGCGCTG 2877  
 DB 1270 GCGCGCTTCGCGCGCGCGGTGCTCACCGCGCGCGCGCTGCTGCTGCTCCCGCGAAC 1329  
 QY 2878 CGTTCAAGTTCACGCGGTTCAGGCGGCGCGGAGGATCTGCTGGGCAAGGACTCGGCG 2937  
 DB 1330 CCGTACAAATACCCGCTGTTCCAGGCGCGGCAACGCGTGTGCTCGGCAAGGACTCGCC 1389  
 QY 2938 TACCTGCAGATGAAGATGGCGCTGGCCATCTCTTGGCTTCTACAGTTCGCGCTGCTG 2997  
 DB 1390 GTACCCGAGATGAAGCGCGGCGGCGCTGCTGAGGGGCTTCGACGCTCGAGTCTGTC 1449  
 QY 2998 GAGGGCA 3005  
 DB 1450 GCGGAA 1457

RESULT 12  
 ACL21306/c  
 ID ACL21306 standard; DNA; 621 BP.  
 XX AC  
 AC ACL21306;  
 XX XX  
 DT 27-OCT-2003 (revised)  
 DT 17-OCT-2003 (first entry)  
 XX XX  
 DE DNA clone originating in barley containing SNP encoding sequence #11297.  
 XX XX  
 KW Barley; single nucleotide polymorphism; SNP; genotype-phenotype analysis;  
 KW gene; ss.  
 XX XX  
 OS Hordeum vulgare; var. (cul.Akashinriki).  
 XX XX  
 FN WO2003057877-A1.  
 XX XX  
 PD 17-JUL-2003.  
 XX XX  
 PF 16-DEC-2002; 2002WO-IB005403.  
 XX XX  
 PR 20-DEC-2001; 2001JP-00387059.  
 PR 20-DEC-2001; 2001JP-00387131.  
 PR 20-DEC-2001; 2001JP-00403299.  
 PR 20-DEC-2001; 2001JP-00403300.  
 PR 27-SEP-2002; 2002JP-00327515.  
 XX XX  
 PA (UYN1-) UNIV JAPAN OKAYAMA.  
 XX XX  
 PI Sato K, Takeda X, Kohara Y;  
 XX XX  
 DR WPI; 2003-587127/55.  
 XX XX  
 PT Single nucleotide polymorphism sites in barley varieties and DNA  
 PT sequences containing them for analysis and identification of barley  
 PT varieties and production of barley transformants with desired  
 PT characteristics.  
 XX XX  
 PS Disclosure; SEQ ID XX; 284pp; Japanese.  
 XX XX  
 CC The present invention relates to oligonucleotide clones originating in  
 CC barley (Hordeum vulgare) which contain single nucleotide polymorphisms  
 CC (SNP). The oligonucleotides may be used for analysis of SNPs among barley  
 CC varieties, identification of particular varieties and genotype-phenotype  
 CC analysis, isolation of specific genes and creation of new varieties by  
 CC transformation of barley varieties with them and production of new barley  
 CC varieties with desired properties. The present sequence represents an

XX	Disclosure; SEQ ID NO 1380; 299pp; English.
XX	
CC	The invention relates to a novel isolated polynucleotide comprising a plant nucleotide sequence having an open reading frame that encodes a polypeptide associated with disease resistance or its fragment having substantially the same activity as the full-length polypeptide. The polynucleotide of the invention is useful for conferring resistance or tolerance to a plant pathogen. The present sequence represents a gene conferring disease resistance used in the invention.
XX	
SQ	Sequence 528 BP; 75 A; 191 C; 183 G; 79 T; 0 U; 0 Other;
	Query Match            2.7%; Score 105.4; DB 8; Length 528;
	Best Local Similarity 61.2%; Pred. No. 2e-13;
	Matches 188; Conservative 0; Mismatches 116; Indels 3; Gaps 1;
QY	2698 GACCCCAAGGGGATCTTGAGGACGACGTGTCGCGGACGGGACGAAAGTGAAGGCCGCGC 2757 
Db	124 GACTCCAAGTTCTGCTGTGGGCGCACAGCGTGCTCCCACGCGGCACCTACGTCGCGGGCGGC 183 
QY	2758 GGGATGGTAGCTACGTGCGCTACTCGATGGGCGCATGGAGTACAACCTGGGCGCCCCGAC 2817 
Db	184 ACGCGGTGACCTACCACCCCTACGCATGGGCGCATGCCGGCGATCTGGGCGCGCAC 243 
QY	2818 GGCGCGAGCTTCGCGCGGAGCGGTGGATCAACGAGAGTGGG--CGTTCCGCAACGCG 2874 
Db	244 CACGGGCGCTTCGCGCGGCGCGCTGGCTCACCGGCGCGCGCGGTGTTGTTCTCCCGAG 303 
QY	2875 TCGCCGTTCAGATTACGGGTTCCAGGGGGGGCCGAGGATCTGCTGGGCAAGGACTCG 2934 
Db	304 AGCCTGTACAAGTACCGGTGTTCCAGCGGGGCTTCGCGTGTGCTTCGCAAGAGCTC 363 
QY	2935 GCGTACCTCGAGATGAAGATGGCGCTGGGCAATCCTCTTGCGCTTCTACAGCTTCGCGCTG 2994 

Qy	2995	CTGGAGG	3001
Db	424	GTCGGGG	430
RESULT 14			
ACL21305/c			
ID	ACL21305	standard; DNA; 399 BP.	
XX	XX	AC	
XX	ACL21305;		
XX	XX		
DT	27-OCT-2003	(revised)	
DT	17-OCT-2003	(first entry)	
XX	XX	DNA clone originating in barley containing SNP encoding sequence #11296.	
DE	Barley;	single nucleotide polymorphism; SNP; genotype-phenotype analysis;	
XX	gene; ss.		
KW	Hordeum vulgare; var. (cul.Akashinriki).		
XX	OS		
XX	WO2003057877-A1.		
PN	17-JUL-2003.		
XX	XX		
PD	16-DEC-2002; 2002WO-IB005403.		
PD	XX		
XX	XX		
PF	20-DEC-2001; 2001JP-00387059.		
XX	PR		
PR	20-DEC-2001; 2001JP-00387131.		
PR	20-DEC-2001; 2001JP-00403299.		
PR	20-DEC-2001; 2001JP-00403300.		
PR	27-SEP-2002; 2002JP-00327515.		
XX	XX		
PA	(UYNI-) UNIV JAPAN OKAYAMA.		
XX	XX		
PI	Sato K, Takeda K, Kohara Y;		

```

XX DR WPI; 2003-587127/55.
XX PF
XX PR
XX PR Single nucleotide polymorphism sites in barley varieties and DNA
XX PT sequences containing them for analysis and identification of barley
XX PT varieties and production of barley transformants with desired
XX PT characteristics.
XX PS
XX PS Disclosure; SEQ ID XX; 284pp; Japanese.
XX PA
XX CC The present invention relates to oligonucleotide clones originating in
XX CC barley (Hordeum vulgare) which contain single nucleotide polymorphisms
XX CC (SNP). The oligonucleotides may be used for analysis of SNPs among barley
XX CC varieties, identification of particular varieties and genotype-phenotype
XX CC analysis, isolation of specific genes and creation of new varieties by
XX CC transformation of barley varieties with them and production of new barley
XX CC varieties with desired properties. The present sequence represents an
XX CC oligonucleotide clone DNA sequence featured in the specification. The
XX CC sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published-pct-sequences. (Updated on 27-OCT-2003 to
XX CC standardise OS field)
XX SQ Sequence 399 BP; 74 A; 162 C; 86 G; 77 T; 0 U; 0 Other;

Query Match 2.7%; Score 104; DB 8; Length 399;
Best Local Similarity 59.5%; Pred. No. 3.8e-13;
Matches 195; Conservative 0; Mismatches 130; Indels 3; Gaps 1;

QY 2741 CGAAGTGAGGCGCGGGGATGGTACGTACGTCCCTACTCGATGGGGGATGGAGT 2800
Db |||||
QY 398 CGCGGTGGGACGGGTGGTTCGTGGCGTACAACTCGATGCGGGGATGGAGT 339
Db |||||
QY 2801 ACAACTGGGGCCCGACCGCGGAGCTTCGGGCCGGAGCGGTGGATCAAC---GAGGATG 2857
Db |||||
QY 338 CTGTGTGGGGGAGACGCGAGGCGGTACCGGCCGGAGCGGTGGACCCGGCGGAGG 279
Db |||||
QY 2858 GCGGTTCCGCAACGCGTCGCGTCAAGTTACGCGGTTCCAGCGGGCGGAGTCT 2917
Db |||||
QY 278 GGAAGTTCGGCGGAGAGCGGTTCGGGTACATGGCGTTTCACCGGGGCCAAGATTT 219
Db |||||
QY 2918 GCCTGGCAAGACTCGCGTACCTGCGATGAAGATGGCGTGGCCATCTCTTGGCGCT 2977
Db |||||
QY 218 GCCTCGAAGAGATGGCGTATATCCAGATGAAGTCTATCGTGGCGTGGTGGAGG 159
Db |||||
QY 2978 TCTACAGCTTCGGCTGCTGGAGGGGACCCGGTGCAGTACCGCATGATGACCATCTCT 3037
Db |||||
QY 158 AGTTTGAATTGGCGTGGACCGCGGTACCGGCCCGGAGGTGGCATCTCTGACGTTAC 99
Db |||||
QY 3038 CCATGGCGCAGGCTCAAGGTCGCGGT 3065
Db |||||
QY 98 GGATGGCGGACGGGCTCCCGTGGCGGT 71
Db |||||

RESULT 15
ID AC121309/c
XX AC121309 standard; DNA; 460 BP.
XX AC AC121309;
XX AC
XX DT 27-OCT-2003 (revised)
XX DT 17-OCT-2003 (first entry)
XX DE
XX DE DNA clone originating in barley containing SNP encoding sequence #11300.
XX KW Barley; single nucleotide polymorphism; SNP; genotype-phenotype analysis;
XX KW gene; ss.
XX OS Hordeum vulgare; var. (cul.Akashinriki).
XX PN WO2003057877-A1.
XX XX
XX PD 17-JUL-2003.

```

```

XX PF 16-DEC-2002; 2002WO-IB005403.
XX PR
XX PR 20-DEC-2001; 2001JP-00387059.
XX PR 20-DEC-2001; 2001JP-00387131.
XX PR 20-DEC-2001; 2001JP-00403299.
XX PR 20-DEC-2001; 2001JP-00403300.
XX PR 27-SEP-2002; 2002JP-00327515.
XX PA
XX PA (UYN1-) UNIV JAPAN OKAYAMA.
XX PI Sato K, Takeda K, Kohara Y;
XX PI WPI; 2003-587127/55.
XX PT Single nucleotide polymorphism sites in barley varieties and DNA
XX PT sequences containing them for analysis and identification of barley
XX PT varieties and production of barley transformants with desired
XX PT characteristics.
XX PS
XX PS Disclosure; SEQ ID XX; 284pp; Japanese.
XX CC The present invention relates to oligonucleotide clones originating in
XX CC barley (Hordeum vulgare) which contain single nucleotide polymorphisms
XX CC (SNP). The oligonucleotides may be used for analysis of SNPs among barley
XX CC varieties, identification of particular varieties and genotype-phenotype
XX CC analysis, isolation of specific genes and creation of new varieties by
XX CC transformation of barley varieties with them and production of new barley
XX CC varieties with desired properties. The present sequence represents an
XX CC oligonucleotide clone DNA sequence featured in the specification. The
XX CC sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published-pct-sequences. (Updated on 27-OCT-2003 to
XX CC standardise OS field)
XX SQ Sequence 460 BP; 95 A; 174 C; 98 G; 91 T; 0 U; 2 Other;

Query Match 2.7%; Score 104; DB 8; Length 460;
Best Local Similarity 59.5%; Pred. No. 3.8e-13;
Matches 195; Conservative 0; Mismatches 130; Indels 3; Gaps 1;

QY 2741 CGAAGTGAGGCGCGGGGATGGTACGTACGTCCCTACTCGATGGGGGATGGAGT 2800
Db |||||
QY 459 CGCGGTGGGACGGGTGGTTCGTGGCGTACAACTCGATGCGGGGATGGAGT 400
Db |||||
QY 2801 ACAACTGGGGCCCGACCGCGGAGCTTCGGGCCGGAGCGGTGGATCAAC---GAGGATG 2857
Db |||||
QY 399 CTGTGTGGGGGAGACGCGAGGCGGTACCGGCCGGAGCGGTGGACCCGGCGGAGG 340
Db |||||
QY 2858 GCGGTTCCGCAACGCGTCGCGTCAAGTTACGCGGTTCCAGCGGGCGGAGTCT 2917
Db |||||
QY 339 GGAGTTCGGCGGAGAGCGGTTCGGGTACATGGCGTTTCACCGGGGCCAAGATTT 280
Db |||||
QY 2918 GCCTGGCAAGACTCGCGTACCTGCGATGAAGATGGCGTGGCCATCTCTTGGCGCT 2977
Db |||||
QY 279 GCCTCGAAGAGATGGCGTATATCCAGATGAAGTCTATCGTGGCGTGGTGGAGG 220
Db |||||
QY 2978 TCTACAGCTTCGGCTGCTGGAGGGGACCCGGTGCAGTACCGCATGATGACCATCTCT 3037
Db |||||
QY 219 AGTTTGAATTGGCGTGGACCGCGGTACCGGCCCGGAGGTGGCATCTCTGACGTTAC 160
Db |||||
QY 3038 CCATGGCGCAGGCTCAAGGTCGCGGT 3065
Db |||||
QY 159 GGATGGCGGACGGGCTCCCGTGGCGGT 132
Db |||||

Search completed: March 7, 2004, 12:36:24
Job time : 968 secs

```

Pred. No. is the number of results predicted by chance to have a

RESULT 1					
AX453578					
LOCUS					
Sequence 7 from Patent WO226789.					
DEFINITION					
AX453578					
ACCESSION					
VERSION					
AX453578.1					
GI:21712820					
KEYWORDS					
.					
Zea mays					
SOURCE					
ORGANISM					
Zea mays					
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;					
Eukarya; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD					
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD					
clade; Panicoidae; Andropogoneae; Zea.					
REFERENCE					
1					
Albertsen,M.C., Fox,T., Huffman,G. and Trimnell,M.					
AUTHORS					
Nucleotide sequences mediating male fertility and method of using					
TITLE					

same		Patent: WO 0226789-A 7 04-APR-2002;		(US)	
JOURNAL		PIONEER HI-BRED INTERNATIONAL, INC.		(US)	
FEATURES		Location/Qualifiers			
source		1..3897			
		/organism="Zea mays"			
		/mol_type="unassigned DNA"			
		/db_xref="taxon:4577"			
ORIGIN					
Query Match		100.0%; Score 3897; DB 6; Length 3897;			
Best Local Similarity		100.0%; Pred. No. 0;			
Matches 3897; Conservative		0; Mismatches 0; Indels 0; Gaps 0;			
QY	1	GAATTCGAAGCGGCCCTTTGACGAGAGTGTGTGATGCGAGTTCGGCGGAATGAGT 60			
DB	1	GAATTCGAAGCGGCCCTTTGACGAGAGTGTGTGATGCGAGTTCGGCGGAATGAGT 60			
QY	61	CGTGCTCAGAGCAACGCTGAGGGTTCCAGGGATGGCAATGGCTATGGCAATCGGCTAG 120			
DB	61	CGTGCTCAGAGCAACGCTGAGGGTTCCAGGGATGGCAATGGCTATGGCAATCGGCTAG 120			
QY	121	AGGTGGAGCAAGGTGCTGAGGATTTGGAGGGCAACCTATGGCAAGTGTGTGAAGGCG 180			
DB	121	AGGTGGAGCAAGGTGCTGAGGATTTGGAGGGCAACCTATGGCAAGTGTGTGAAGGCG 180			
QY	181	ACGCAATGAGAGATCTATTACAGCTTACACTGGATGCCGCCCAACAAATTCACCTTTAGA 240			
DB	181	ACGCAATGAGAGATCTATTACAGCTTACACTGGATGCCGCCCAACAAATTCACCTTTAGA 240			
QY	241	TTTTGATACCTGCTACCTACTTTATTCTTGGTTGGCAACTTCCAAATAGGCTCATGTT 300			
DB	241	TTTTGATACCTGCTACCTACTTTATTCTTGGTTGGCAACTTCCAAATAGGCTCATGTT 300			
QY	301	AATCAATGATGATGATTTATTCAGCAATATTTCTTGTGTTGATGATTTAAATATG 360			
DB	301	AATCAATGATGATGATTTATTCAGCAATATTTCTTGTGTTGATGATTTAAATATG 360			
QY	361	GGTGAGCGGATTAATATCATCATGAGAGCTTTATCTTCAATGCTCTCTTGATTTGG 420			
DB	361	GGTGAGCGGATTAATATCATCATGAGAGCTTTATCTTCAATGCTCTCTTGATTTGG 420			
QY	421	TTTTCAGATCATTTCTTCACTGTTTCAAGAAATTTCTCAGTTTGGTCCATGATTTTG 480			
DB	421	TTTTCAGATCATTTCTTCACTGTTTCAAGAAATTTCTCAGTTTGGTCCATGATTTTG 480			
QY	481	AAGTGAAGTTCCCTTAAATTTCAATATGCTTCTCTTCTTCTTCTTCTTCTTCTTCT 540			
DB	481	AAGTGAAGTTCCCTTAAATTTCAATATGCTTCTCTTCTTCTTCTTCTTCTTCTTCT 540			
QY	541	CTTTTCACTTTGGGTTTCAAAATTTGACTCACAAGAAACAAATTTCACTTTTGGGTTTCA 600			
DB	541	CTTTTCACTTTGGGTTTCAAAATTTGACTCACAAGAAACAAATTTCACTTTTGGGTTTCA 600			
QY	601	AATTCCTCTTCAGGATGATCTTTTCACTTTGATCTGATGATGATGATGATGATGATG 660			
DB	601	AATTCCTCTTCAGGATGATCTTTTCACTTTGATCTGATGATGATGATGATGATGATG 660			
QY	661	CAGTTTTTAAGAACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 720			
DB	661	CAGTTTTTAAGAACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 720			
QY	721	CAGACTTTTGTACCAAGCTGATGATGATGATGATGATGATGATGATGATGATGATG 780			
DB	721	CAGACTTTTGTACCAAGCTGATGATGATGATGATGATGATGATGATGATGATGATG 780			
QY	781	ACTGGCAACTCTTAATTTGATAATAAAGAAATATAATATAGATATCAGATATCTCAT 840			
DB	781	ACTGGCAACTCTTAATTTGATAATAAAGAAATATAATATAGATATCAGATATCTCAT 840			
QY	841	TTGGTTGGCAGATCACAAGAAAGAACAAAGGCTAAGGCTTACTTCTTCTTCTTCTTCT 900			
DB	841	TTGGTTGGCAGATCACAAGAAAGAACAAAGGCTAAGGCTTACTTCTTCTTCTTCTTCT 900			

QY	901	GGTCAAGGACACATATGAATGAAGAAATCTTAATTTGGGGTCACACCAAGATTGTCTC	960
DB	901	GGTCAAGGACACATATGAATGAAGAAATCTTAATTTGGGGTCACACCAAGATTGTCTC	960
QY	961	TTCTCGAGGTTGGGGGTCCTTAAGGTTGGTAGCAATACCCAAATATATCACTTAACAA	1020
DB	961	TTCTCGAGGTTGGGGGTCCTTAAGGTTGGTAGCAATATACCCAAATATATCACTTAACAA	1020
QY	1021	ACCCAATTCATGCTACATACATACATAGCATCCATCTTGTAGCTGGACCTTTCATCA	1080
DB	1021	ACCCAATTCATGCTACATACATACATAGCATCCATCTTGTAGCTGGACCTTTCATCA	1080
QY	1081	AGAGCACCATGGAGGAAGCTCACAATCGCCGGCGAGCCCATGCCATCTTCCCACTAG	1140
DB	1081	AGAGCACCATGGAGGAAGCTCACAATCGCCGGCGAGCCCATGCCATCTTCCCACTAG	1140
QY	1141	CAGGGCCTCACAAGTACATCGGCTCCTCTGTTGTTGCTCTCATGGATCCTGTGTCAGA	1200
DB	1141	CAGGGCCTCACAAGTACATCGGCTCCTCTGTTGTTGCTCTCATGGATCCTGTGTCAGA	1200
QY	1201	GGTGGAGCCTGAGGAAGCAGAAAGCCCGAGATCATGGCCAGTTCATCGGTGCAACGGTGG	1260
DB	1201	GGTGGAGCCTGAGGAAGCAGAAAGCCCGAGATCATGGCCAGTTCATCGGTGCAACGGTGG	1260
QY	1261	AGCAGCTGAGGAAGTACCAACCGGATGCAAGCTGGCTTGTGGGTACTGTCAAGGACA	1320
DB	1261	AGCAGCTGAGGAAGTACCAACCGGATGCAAGCTGGCTTGTGGGTACTGTGTCAAGGACA	1320
QY	1321	GGCAGTGACCGTTCGAGATCGGTTCACTTCTTACACCTACATCGCTGACCCGGTGAATG	1380
DB	1321	GGCAGTGACCGTTCGAGATCGGTTCACTTCTTACACCTACATCGCTGACCCGGTGAATG	1380
QY	1381	TCGAGCATGCTCCTCAAGACTAACTTACCAATTAACCAAGTTAAATGACCTGAACTCAC	1440
DB	1381	TCGAGCATGCTCCTCAAGACTAACTTACCAATTAACCAAGTTAAATGACCTGAACTCAC	1440
QY	1441	TGATGTTGAGTCTTCGGAAATCAGAGCTGAAAGTGAATCGAATGTGCTGAAACCGGTG	1500
DB	1441	TGATGTTGAGTCTTCGGAAATCAGAGCTGAAAGTGAATCGAATGTGCTGAAACCGGTG	1500
QY	1501	TAGGGAATTCGTTGACAGATCCTACATGAGCGTCTCTTCGTTGACGGCATCTTCAACGCC	1560
DB	1501	TAGGGAATTCGTTGACAGATCCTACATGAGCGTCTCTTCGTTGACGGCATCTTCAACGCC	1560
QY	1561	GACGGCGAGCTGTGAGGAAGCAGAGGAAGACGGCGAGTTTCGAGTTCCCTTCCAGAAC	1620
DB	1561	GACGGCGAGCTGTGAGGAAGCAGAGGAAGACGGCGAGTTTCGAGTTCCCTTCCAGAAC	1620
QY	1621	CTGAGGATTTTCAGCGCATTTGTGTTTCAGAGAGTACTCCCTGAAGCTGTGCGGTATCTG	1680
DB	1621	CTGAGGATTTTCAGCGCATTTGTGTTTCAGAGAGTACTCCCTGAAGCTGTGCGGTATCTG	1680
QY	1681	AGCCAGGCAATCAAGGCGAGGCAAGTTGTGGACATGCAAGTGTGCGGTATCTG	1740
DB	1681	AGCCAGGCAATCAAGGCGAGGCAAGTTGTGGACATGCAAGTGTGCGGTATCTG	1740
QY	1741	CATTGCGAATGAGCATTTCAACCTGAGACACGAGAGTACTCTTGCCTTCCAGAACT	1800
DB	1741	CATTGCGAATGAGCATTTCAACCTGAGACACGAGAGTACTCTTGCCTTCCAGAACT	1800
QY	1801	TTACATGAGATGAGCTGGAATCTCCATCTGCAAGGTTGGGTCGAGATCGGCAC	1860
DB	1801	TTACATGAGATGAGCTGGAATCTCCATCTGCAAGGTTGGGTCGAGATCGGCAC	1860
QY	1861	GCTGTGCGCGGATCTCCCGGAGAACAGCTTCGCGCAGGCGTTTCGATGCGCGCAATCAT	1920
DB	1861	GCTGTGCGCGGATCTCCCGGAGAACAGCTTCGCGCAGGCGTTTCGATGCGCGCAATCAT	1920
QY	1921	CGTCAAGCTGCGGTTTCATCGAAGGCTGTGGCGCATCAAGAGGTTCTTCCAGTGGGTC	1980
DB	1921	CGTCAAGCTGCGGTTTCATCGAAGGCTGTGGCGCATCAAGAGGTTCTTCCAGTGGGTC	1980





Db 1320 GTACGTGCCCTACTCGATGGGGCGGATGGAGTCAACACTGGGGCCCGGACGGCGAGCTT 1379  
Qy 2829 CCGGCGGAGCGGTGGATCAACGAGGATGGCGCGTTCGCCAACGCGTCCCGCTTCAAGTT 2888  
Db 1380 CCGGCGGAGCGGTGGATCAACGAGGATGGCGCGTTCGCCAACGCGTCCCGCTTCAAGTT 1439  
Qy 2889 CACGCGCTTCAGGGCGGCGGAGGATCTGCTGGGCAAGGACTCGGCTACCTGTCAGAT 2948  
Db 1440 CACGCGCTTCAGGGCGGCGGAGGATCTGCTGGGCAAGGACTCGGCTACCTGTCAGAT 1499  
Qy 2949 GAAGATGGCGTGGCATCTCTTCGCTTCTACAGCTTCGCGCTGCTGGAGGGCACCC 3008  
Db 1500 GAAGATGGCGTGGCATCTCTTCGCTTCTACAGCTTCGCGCTGCTGGAGGGCACCC 1559  
Qy 3009 GGTGAGTACCGCATGATGACCATCTCTCCATGGCGGACGCGCTTCAAGGTCGCGTCTC 3068  
Db 1560 GGTGAGTACCGCATGATGACCATCTCTCCATGGCGGACGCGCTTCAAGGTCGCGTCTC 1619  
Qy 3069 TAGGCGCGTCTGATGATGATGCGGATTTG-----GGATATCATCCCGCTTAATCC 3117  
Db 1620 TAGGCGCGTCTGATGATGATGCGGATTTGATGATGATGATGATGATGATGATGATGAT 1679  
Qy 3118 -----TTAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3164  
Db 1680 AATAAGCTCGTGTACAAATTTGATGATGATGATGATGATGATGATGATGATGATGAT 1739  
Qy 3165 GGTGCTTGGCTTAAGCCCTTAAATTTGATGATGATGATGATGATGATGATGATGATGATGAT 3224  
Db 1740 GGTGCTTGGCTTAAGCCCTTAAATTTGATGATGATGATGATGATGATGATGATGATGATGAT 1799  
Qy 3225 GTTTTGTACTCTACTCTCAGTGGAGTGTAGTGCAGCATCAAGTTCATCATATATAT 3284  
Db 1800 GTTTTGTACTCTACTCTCAGTGGAGTGTAGTGCAGCATCAAGTTCATCATATATAT 1859  
Qy 3285 TATCCTCTTCTT 3297  
Db 1860 TATCCTCTTCTT 1872

RESULT 3  
AX453572  
LOCUS AX453572 1906 bp DNA linear PAT 02-SEP-2002  
DEFINITION Sequence 1 from Patent WO0226789.  
ACCESSION AX453572  
VERSION AX453572.1 GI:21712815  
KEYWORDS  
SOURCE Zea mays  
ORGANISM Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.  
1  
REFERENCE Albertsen, M.C., Fox, T., Huffman, G. and Trimmell, M.  
AUTHORS Nucleotide sequences mediating male fertility and method of using  
TITLE same  
JOURNAL Patent: WO 0226789-A 1 04-APR-2002;  
PIONEER HI-BRED INTERNATIONAL, INC. (US)  
FEATURES Location/Qualifiers  
source 1. .1906  
/organism="Zea mays"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:4577"  
join(.1.1638,1642..1770)  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAD38131.1"  
/db\_xref="GI:21712816"  
/db\_xref="EMBL:CAD38131"  
/translation="EFQREAHITPATPSPFFFLAGHKYIALLLVVLVILVQVMSL  
RKQGRSPWPIGATVEQLRNHRMDLWYLSLRHTVDMPTFSYTIADPVAIVE  
HVLKTNFTNPKGIVYRSMVLLDGIIFNADGELWRKORTASFFASKNLRFSAI  
VREYSLSLQSLIISQASKAKVDMQELNRMITLDSICKVGFVETGLSPDLPENSF  
AQAFDANIITLRFIDPLNRIKRFHFVHSGEALLAQSIKLVDFTSYVIRRRKAEIVE

VRASGKQKMKHDLISRFIELGAGDDGGGDDKSLRDLVNLVFIAGRTDTATLSW  
FTWAKSHPOVAEKLEELCAFEAREARESEVTLVLCGADADDKAPAAVAQPAFL  
TYDSLKGLVYLHACVTETLRYPAVPQDPKGLLEDDVLPGTKVYRAGMTVYVYVSMG  
RMEYNWPDAAASFRPERKINEDAFNAPSPKFTAFQAPRILCKGDSAYLQMKMALA  
ILFRFYSFLLEGPVQYRMNTILSMHAGLKVRSRVACHGDLDMDIVLNPRLITLV  
LQICMHACKGRWVSLVAVLKP"  
ORIGIN  
Query Match 35.4%; Score 1378.4; DB 6; Length 1906;  
Best Local Similarity 82.1%; Pred. No. 3e-198;  
Matches 1833; Conservative 0; Mismatches 16; Indels 385; Gaps 6;  
Qy 1088 CATGAGGAACTCACATCACGCCGGGACGCCATCGCCATTCCTCCACTAGCAGGCCC 1147  
Db 9 CACGAGGAACTCACCTCACGCCGGGACGCCATCGCCATTCCTCCACTAGCAGGCCC 68  
Qy 1148 TCACAAGTACATCGCGTCTCTTGGTTGTCTCTCATGATTCCTGTGTCAGAGTGGAG 1207  
Db 69 TCACAAGTACATCGCGTCTCTTGGTTGTCTCTCATGATTCCTGTGTCAGAGTGGAG 128  
Qy 1208 CCGTGGGAGCAGAAAGCCGAGATCATGCCAGTCTCGGTGACGAGTCAACGCTGGAGCAGCT 1267  
Db 129 CCGTGGGAGCAGAAAGCCGAGATCATGCCAGTCTCGGTGACGAGTCAACGCTGGAGCAGCT 188  
Qy 1268 GAGGAATACCAACCGGATGCACGACTGGCTTGTTCGGGTACCTGTCCAGGCAACAGACAGT 1327  
Db 189 GAGGAATACCAACCGGATGCACGACTGGCTTGTTCGGGTACCTGTCCAGGCAACAGACAGT 248  
Qy 1328 GACCGTCAACATGCGGTTCATCTTCTACCTACCTACCTGACCGGTGATGTCAGCA 1387  
Db 249 GACCGTCAACATGCGGTTCATCTTCTACCTACCTGACCGGTGATGTCAGCA 308  
Qy 1388 TGTCTCTCAAGTACTTACCAATTTACCCAAAGTAAATGACCTGAATCACTCATGATGT 1447  
Db 309 TGTCTCTCAAGTACTTACCAATTTACCCAAAGTAAATGACCTGAATCACTCATGATGT 340  
Qy 1448 CAGTCTTCGGAATCAGAGCTGAAAGTGAATCGAATGTGCTGAACACCGTGTAGGAA 1507  
Db 341 -----AGGGA 346  
Qy 1508 TCGTGTACAGATCTTACATGAGACGTGCTCTCGGTGACGGCATCTTAAACGCGGAGGCG 1567  
Db 347 TCGTGTACAGATCTTACATGAGACGTGCTCTCGGTGACGGCATCTTAAACGCGGAGGCG 406  
Qy 1568 ACTGTGAGGAGACAGAGGAGAGCGGAGTTGAGTTCGCTCCCAAGACCTGAGG 1627  
Db 407 AGCTGTGAGGAGACAGAGGAGAGCGGAGTTGAGTTCGCTCCCAAGACCTGAGG 466  
Qy 1628 ATTTGAGGCGCCATTTGTTTCAGAGAGTACTCCCTGAAGCTGTGGGTATATGAGCCAG 1687  
Db 467 ATTTGAGGCGCCATTTGTTTCAGAGAGTACTCCCTGAAGCTGTGGGTATATGAGCCAG 526  
Qy 1688 ATTCGAAGCAGGAAAGTTGGACATGACAGTGCAGTGCAGTGCCTCCCTGCAATGCC 1747  
Db 527 CATCCAAGCAGGAAAGTTGGACATG----- 555  
Qy 1748 ACATGAGCATTTCAACCTGAGACACAGAGCTACCTTCCGATTCAGGAACCTTTACATG 1807  
Db 556 -----CAGGAACCTTTACATG 570  
Qy 1808 AGGATGACGCTGACCTCCATCTGCAAGGTTGGGTTCGGGTTCGAGATCGGACGCTGTGCG 1867  
Db 571 AGGATGACGCTGACCTCCATCTGCAAGGTTGGGTTCGGGTTCGAGATCGGACGCTGTGCG 630  
Qy 1868 CCGGATCTCCCGAGAACAGTTCGCGGAGGCTTCGATGCGGCCCAACATCATCTGTCAG 1927  
Db 631 CCAGATCTCCCGAGAACAGTTCGCGGAGGCTTCGATGCGGCCCAACATCATCTGTCAG 690  
Qy 1928 CTGCGGTTTCATGACCGCTGTGGCGCATCAGAGGTTCTTCCACGTCGGTTCAGAGGCC 1987  
Db 691 CTGCGGTTTCATGACCGCTGTGGCGCATCAGAGGTTCTTCCACGTCGGTTCAGAGGCC 750  
Qy 1988 CTCTAGCGCAGAGCATCAAGCTCGTGGACGAGTTTCCCTACAGCGTGTATCCCGCGAGG 2047



genscan and Genscan+ (Chris Burge, <http://CCR-081.mit.edu/GENSCAN.html>), GeneMarkEM (Mark Borodovsky, <http://genemark.biology.gatech.edu/GeneMark/>), and GeneSplicer (Mihaela Pertea and Steven Salzberg, contact [mpertea@tigr.org](mailto:mpertea@tigr.org)), searches of the complete sequence against a peptide database and the plant EST database at tigr (<http://www.tigr.org/tdb/tgi.shtml>). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are annotated by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by RepeatMasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>).

## FEATURES

source

1. .160738

/organism="Oryza sativa (japonica cultivar-group)"

/mol\_type="genomic DNA"

/cultivar="Nipponbare"

/db\_xref="taxon:39947"

/chromosome="3"

/map="SL13095"

/clone="OSJNBa0091P11"

/notes="japonica cultivar-group"

1. .2006

/gene="OSJNBa0091P11.1"

/note="near identical to sucrose transporter (C terminus)"

GB:BAA24071 GI:2723471 (Oryza sativa) (Plant Cell Physiol

38 (12), 1389-1396 (1997)); EST C27284, D40515, A0108830

from this gene"

join(&lt;1. .41,267. .461,580. .723,833. .1066,1169. .1232,

1316. .1359,1476. .2006)

/gene="OSJNBa0091P11.1"

join(&lt;3. .41,267. .461,580. .723,833. .1066,1169. .1232,

1316. .1359,1476. .1649)

/gene="OSJNBa0091P11.1"

/codon\_start=1

/product="sucrose transporter, 5'-partial"

/protein\_id="AAL84308.1"

/db\_xref="GI:19071641"

/translation="ACANLKAGFLVAVIFSLCLVTLIFAKEVPPKGNALPTKNE

PAPECTGLAVLKGPRNPTGMPSVLIVTGLTWSFFPILYDTDMGREIVHGDPK

GDPOQTEAFNOGVRAAGAFGLLINSIVLGFSSFLIEPMCKVGRVVMVTSFLVCIAM

AATPLSFLSLKDFHTVOKATPADKSIKAVCLVLEFLGVPLAVLYSVFVATOLA

ATFGGQGLCTGVNLIVTPOVVIAGAGPWDELFGKGNIPAFGLASGFALLGVAG

IFLLPKIKRQKFRSVSMGGH"

1662. .1714

/rpt\_family="(CGG)n"

complement(2364. .2386)

/rpt\_family="(GA)n"

complement(3213. .3250)

/rpt\_family="AT-rich"

complement(3415. .3448)

/rpt\_family="AT-rich"

complement(4151. .5119)

/gene="OSJNBa0091P11.1"

/notes="predicted by fgenesh"

complement(&lt;4151. .5119)

/gene="OSJNBa0091P11.1"

complement(4151. .5119)

/gene="OSJNBa0091P11.1"

/codon\_start=1

/product="hypothetical protein"

/protein\_id="AAL84309.1"

/db\_xref="GI:19071642"

/translation="MGCTTSRQARHLRHCFSPALPRCOSFPARCGADGAVHVRLT

STTLGSLVDKGAFAAEAFMRVPTMTPTNPEPAIDAWAGLEESPLV

PPFAHGSFPTTAVPELAASRKVTPLVLEKKKASPVAPRKAVLYFTSIRGVRA

THECCCLARALGGVGVVERDYSMERGFDELHGLGRGAALAKCWAPAAPAL

PSLFDVDELGVNDELKELHLAGELARLAGCESAAPGEAGACEACADVRFVLCGAC

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

SGSKYVVDGDDDDENPLDGGGGGPRCTECNENGIIVRCPVCC"

complement(4245. .4267)

/rpt\_family="(CGA)n"

complement(4337. .4365)

/rpt\_family="GC-rich"

complement(4468. .4491)

/rpt\_family="GC-rich"

complement(4925. .4955)

/rpt\_family="GC-rich"

complement(7403. .7423)

/rpt\_family="AT-rich"

complement(7998. .8021)

/rpt\_family="AT-rich"

8348. .8398

/rpt\_family="(AGGGGG)n"

complement(8923. .8951)

/rpt\_family="(GA)n"

complement(8983. .9010)

/rpt\_family="AT-rich"

9203. .9268

/rpt\_family="tandem"

11311. .12015

/gene="OSJNBa0091P11.18"

/note="predicted by fgenesh"

join(&lt;11311. .11749,11879. .12015)

/gene="OSJNBa0091P11.18"

join(11311. .11743,11879. .12015)

/gene="OSJNBa0091P11.18"

/codon\_start=1

/product="hypothetical protein"

/protein\_id="AAL84310.1"

/db\_xref="GI:13071643"

/translation="MGGLGVLDLDFARALRLWLFWECPCDKPWGTTPHYDELDE

QIFVAATVTIGNGAKAFWTSNWIHQPLKILAPALNHSKGRQLVQDALTDNKKI

EDIRHLSMPLIRFEFAVRLWNETNLEEGVEDITITWRWTNTGIGHPTACNSDNGP

ITTFANSATGTVKRNYSRIVHSREKGTG"

complement(12707. .12726)

/rpt\_family="(GA)n"

12726. .12754

/rpt\_family="tandem"

12831. .15245

/gene="OSJNBa0091P11.3"

/note="similar to DNA binding protein GB:BAA05625

GI:1435022 (Daucus carota); EST D48120 from this gene"

join(&lt;12831. .13484,14329. .15245)

/gene="OSJNBa0091P11.3"

join(12831. .13484,14329. .14775)

/gene="OSJNBa0091P11.3"

/codon\_start=1

/product="putative DNA-binding protein"

/protein\_id="AAL84311.1"

/db\_xref="GI:19071644"

/translation="MASNGMASPSPFFPNFLHMAQQQAAPPDPQEHHHHHHHHGH

HGHHEQQQQQHHHHLGPPPPPPHPHPNPFPLSSAQCSLQFFRGMAPLGRPMYSY

GDGGGGDEVGNGDELSDGSGQAGEKKRLNVEQVTLKFNGLKLEPRKQVL

ARALGLQPRQVAFQNRARWTKOLEXDYDLKQLDAKDAENDALHNKKLQAE

IYALKGREASSELINLNETEASCNSRNSSEINLIDISRTTPDAAALDTAPTAGHH

HGGGGGGGGGGMIPFYYSIARPSAGSGGVVDIDQLHSSSGGAGGPKMEHGGGNNVQ

AASVPTASFNLLCGPEPPFPFPWPDHGFH"

complement(12936. .13011)

/rpt\_family="(TGG)n"

13136. .13179

/rpt\_family="(CGG)n"

complement(13641. .13662)

/rpt\_family="(CAG)n"

13848. .13884

/rpt\_family="(CGAT)n"

14509. .14541

/rpt\_family="tandem"

complement(15170. .15194)

/rpt\_family="AT-rich"

complement(16201. .16222)

/rpt\_family="(TA)n"

repeat_region	complement (18417..18442)	
Query Match 28.6%; Score 1114.8; DB 8; Length 160738; Best Local Similarity 73.3%; Pred. No. 1.3e-158; Matches 1671; Conservative 0; Mismatches 432; Indels 177; Gaps 12;		
Qy	836	TTTTCTTGGTGGCAGATCAAAAAGGAGACAAAAGGCTAAGCCTCTACTGTTTCGGG 895
Db	95440	TTTTCTTAGGTGACAGATCACAAAAGGAAACTAAAG--CTAAGATCCAATCTCTTAAGG 95497
Qy	896	AGTTAGGTCAAGGACACCATATGAATGAAAGAAATCTTTAAATTTGGGGTCACACCAAGATT 955
Db	95498	TGTTAGGTTAGGACACCATATGAATGAGA-CAATCTTTAAATTTCTTGGTCAACAAGATT 95556
Qy	956	GTCTCTCTCGAGTTGGGGGGTCCCTAAGGTTGGTAGCAATACCAATACCAATATATCACT 1015
Db	95557	GTCTC-----AAAGTTGGTAGCATCAGTCCCAATATATACCT 95595
Qy	1016	AACAAACCAATCCATGCTACATACATATAGCATCCATCACTTGTAGACTGGACCCCTT 1075
Db	95596	AA-----CTATGCCATCCAAAATGCTACATAGCATCTCTTGTAGACTGAACCCCTT 95645
Qy	1076	CATCAAGACACCATGGAGGAAGCTCACATCACGCGCGGCGAGCCCATCGCCATCTTCCC 1135
Db	95646	CATGAAGAGCCCATGGAGGAGCTCATGCAATGCCAGTGA-----CATCATTTCTTCCC 95699
Qy	1136	ACTAGCAGGCGCTCAAGATACATCGCGCTCCTCTGTTGTCCTCTCATGATCCTGGT 1195
Db	95700	AGTAGCAGGAATCCAAGGCTCATAGCTATCTTCTTGTGTTGTCCTCTCATGGATCTTGGT 95759
Qy	1196	CCAGAGTGGAGCCTGAGGAAGCAGAAAGCCCGAGATCATGGCCAGTCATCGGTGCAAC 1255
Db	95760	CCACAAGTGGAGCTGAGGAACCCAGAAAGGCGCAAGATCATGGCCAAATCATCGGCGCAC 95819
Qy	1256	GGTGGAGAGCTGAGGAACCTACACCGGATGACAGACTGGCTTGTCCGTACCTGTACG 1315
Db	95820	AGTGGAGCAACTGAAGAACTTACCACAGGATGATGACTGGCTGTGCGAGTACTTGTGAA 95879
Qy	1316	GCACAGACAGTGACCGTGACATGCGGTTTCACTTCTTACACTACATCGCTGACCCGGT 1375
Db	95880	GGACAGGACGGTGACCGTGACATGCTTTTCACTTCTTACCTCTTACACTATTTGCCGACCCGGT 95939
Qy	1376	GAATGTGAGCATGTCTCTCAAGACTAATCTTCACTAAATTAACCCCAAGGTAATGAOCTGAA 1435
Db	95940	GAACGTCGACATGTCTCTGAAGACCAACTTACCAATTTACCCCAAGGTAAGAAACCAATA 95999
Qy	1436	CTCACTGATGTTTCACTTCTCGGAAATCAGAGC-----TGAAGCTGAATCGAATGTGCC 1489
Db	96000	GGATCTTCAGTGACTGTAAATGTGCTTGGACAGTACTAACACTGACACAAAATGT 96059
Qy	1490	TGAACACCGGTGAGGAATCGGTACAGATCCTTACATGACGCTGCTCCTCGGTGACGCA 1549
Db	96060	CTGAAATAATGACGGGTGAAGTGTACAGGTCTTACATGGATGTGCTGCTCGGTGATGCA 96119
Qy	1550	TCTTCAAGCCGACGGGAGCTGTGAGGAGACAGAGGAGCGGCGAGTTTCGAGTTCCG 1609
Db	96120	TATTCATGCCGACGCGAGATGTGAGGAAGCAAAAGGAGACGGCGAGCTTCGAGTTTG 96179
Qy	1610	CCTCCAAAGAACCTGAGGGAATTCAGCGCCATTTGTGTTTCAGAGAGTACTCCCTGAAGCTGT 1669
Db	96180	CCTCCAAAGAACCTGAGAGACTTTCAGCACTGTGTGTTCAGGAGTACTCCCTGAAGCTAT 96239
Qy	1670	CGGGTATCTGAGCCAGGATCCAAAGGACGAGCAAAAGTTTGGACATGCGAGGTGAGATCAC 1729
Db	96240	CAAGCACTCTGAGCCAAAGCATGCAAGCCGCGAGAGTTGTAGACATGCAAGTAACCAACT 96299
Qy	1730	TGCTCCCTTGGCATTGCACATGAGCATTTCAACCTGAGACACAGAGAGCTACCTTGGCG 1789
Db	96300	GAATTCCTTGCCTTAATACCTTAACATTTCTTGAGAAACCAAAATGTTTCAGAAATTCGTATG 96359
Qy	1790	-----ATTCAAGGAACCTTATAGAGGATGACGCTGGACCTCCATCTGCAAG 1834

96360	CAAGAACTAACAAAATTCAGGAATTTGTTTCATGAGGATGACACTGCGACTCGATCTGCAAG	96419	
Qy	1835	GTTTGGGTTCCGGGTTCAGATTCGGACACGCTGTCCGCGGATCTCCCGAGAACAGCTTCGCG	1894
Db	96420	GTCCGGTTTGGGTTTGAATTCGGATCGGACGCTGTACACTGATCTCCCGGAGAACAGCTTTGCC	96479
Qy	1895	CAGCGCTTCGATCCCGCCAAACATCATCTCAGCTCGGTTTCATCGACCGCTGTGGCGC	1954
Db	96480	CAGGCATTGACGCTGCCAACATCATCTCAGCTCAGCTCGGTTTCATCGATCTCTGTGGCGT	96539
Qy	1955	ATCAAGAGGTTCTTCAAGTCCGGTCAGAGGCCCTCCTAGCGAGAGCATCAAGCTCGTG	2014
Db	96540	CTGAAGAAGTTCTTGAACGTCGGATCAGAGGCTCTCTCTCGAGCAGAGCATGAAGCTGGTT	96599
Qy	2015	GACGAGTTTCACTTACAGCTGATCCGCCGGAGAGGCCGAGATCTGTCGAGGCCCGGGCC	2074
Db	96600	GATGACTTCACTTACAGCTGATCCGCCGCCGCAAGGCTGAGATCTTGCAGGCTCGAGCC	96659
Qy	2075	AGCGGCAACAGGAGAAGGTACGTGCACATGACTGTTTTCGATTCTTTCAGTTTCATCGTCTT	2134
Db	96660	AGCGGCAAGCAAGAAGGTGATCTCTCTCTCTGCTCAAGAAATCAGTAGAACTGAAC	96719
Qy	2135	GGCGGGATGGACTGATCCTGATTGATTATATATCCGTGTGACTTG-----TGAGGACAA	2190
Db	96720	TGACATGTTAATGTTGATGATCAGATCGGAAAAGGTTTGTGTTCTTGTATATCGTTGATT	96779
Qy	2191	ATTAATAATGGGCAGATGAAGCACGACATCTCTGTCACGTTTCATCGAGCTAGGCGAGCGG	2250
Db	96780	GTAATGGCGAGCAGATCAAGCACGACATACTGTCCGGTTTCATCGAGCTCGGGAGGCGG	96839
Qy	2251	GCAGACGAC-----GGCGGCGGCTTCGGGACGACAGAGAGCTTCGGGACGTGTGCTCA	2304
Db	96840	GCGGCGACGAGGGGGCGGAGCTTCGGGACGACAGAGAGCTTCGGGACGTGTGCTCA	96899
Qy	2305	ACTTCGTGATCCGCGCGCGGACACGACGCGGACGACGCTGTCTGTGTTTACGCACTGG	2364
Db	96900	ACTTCGTGATCCGCGCGCTGACACGACGCGGACGAGCTGTCTGTGTTTACGCTACATGG	96959
Qy	2365	CCATGTCCCAACCCGGAAGTGGCCCGAGAGCTGCGCCCGGAGCTGTGCGGCTTCGAGCGG	2424
Db	96960	CGATGACGACCCCGCGCTCGCCACAAAGCTCCGCGCGGAGCTGGCCGCGCTTCAGGATG	97019
Qy	2425	AGCGCGCGCGGAGGGGGTTCGCTCTGTCGCTCGCGCGGCTGACGCGGACGACA	2484
Db	97020	AGCGCGCGCGGAGGGGGTTCGCTCTGTCGCT-----CGCGGACGCGCGCGGAGG	97067
Qy	2485	AGGCGTTCCGCGCGCGGTGGCGAGTTTCGCGGCGCTCTCACTACGACAGCCTCGGCA	2544
Db	97068	CGTCTTTCGCGCGCGGTGGCGAGTTTCGCTCGCTGCTGAGCTACGACGCGGTGGGA	97127
Qy	2545	AGCTGGTCTACCTCAAGCTTCGCTACCGGAGAGCTTCGCGCTGACCCCGCGCTCCCTC	2604
Db	97128	AGCTGGTGTACTCTCAAGCGGTGCGTGACGGAGAGCTTCGCGCTCTACCCGCGGTGC	97184
Qy	2605	AGGTGAGCGCGCGGACACGCGACCTCCGCTCCAGAGCACAGCATGCAAGTGGGACCT	2664
Db	97185	-----	97184
Qy	2665	GAATGCAATGCAATGCACTTTCGCGCGCGGAGAGACCCCAAGGGGATCTCTGAGGACGAC	2724
Db	97185	-----CGCAGGACCCCAAGGGGATCTCTGAGGAGGACGAC	97216
Qy	2725	GTGCTTCGCGGACGAGCAAGGTGAGGGCGGGGAGTGTGAGCTAGCTAGCTGCTCTACTCG	2784
Db	97217	GTGCTTCGCCGACGCAACAGGTGCGCGCGCGCGGAGTGGTGAAGTACGTACGTGCTCTACTCC	97276
Qy	2785	ATGGGCGGATGGAATTAACACTGGGGGCCCGGACGCGGAGCTTTCGCGCCGAGCGGTGG	2844
Db	97277	ATGGGAGAGTGAATTAACACTGGGGCCCCGACGCGGAGCTTTCGCGCCGAGCGGTGG	97336
Qy	2845	ATCA---ACGAGGATGGCGGCTTCGCAACCGCTGCGGTTTCAAGTTTCAAGGCTTCAG	2901
Db	97337	CTCAGCGGACGCGCGCGCTTCGGAACGCGTCTCGGTTTCAAGTTTCAAGGCTTCAG	97396



```
promoter
1..1088
/note="cytochrome P450-like protein; anther specific"
TATA_signal
1006..1009
/gene="ms*-sb200"
/gene="ms*-sb200"
```

promoter 1.1088

1. 1088

```
/gene="ms*-sb200"
```

TATA signal

1006. 1009

```
/gene="ms*-sb200"
```

## ORIGIN

Query Match

Best Local Similarity 100.0%: Pred. No. 7.2e-155:

Matches 1091; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	GAATTC	CAAGCAGG	CCCTTGT	GTACGAGAGT	GTTC	TGATGTCAGT	CGGCGGAAT	GAGT	60
Db	1	GAATTC	CAAGCAGG	CCCTTGT	GTACGAGAGT	GTTC	TGATGTCAGT	CGGCGGAAT	GAGT	60
QY	61	GCGTCT	GTAGAGCA	ACGCTG	AGGGGTTC	CAAGGATG	GCAATGG	CTATGGCAAT	CGGCTAG	120
Db	61	GCGTCT	GTAGAGCA	ACGCTG	AGGGGTTC	CAAGGATG	GCAATGG	CTATGGCAAT	CGGCTAG	120
QY	121	AGGTGG	AGGACAAAG	GTGGT	GAGGATGGG	AGGCAAC	TATGGCAAGT	TGGTGAAGGC	180	
Db	121	AGGTGG	AGGACAAAG	GTGGT	GAGGATGGG	AGGCAAC	TATGGCAAGT	TGGTGAAGGC	180	
QY	181	ACGCAAT	CGAGAGTCT	ATTCAG	ACTTAC	ACTGGATG	CGGCCAA	CAAAATTC	CAACCTTTAGA	240
Db	181	ACGCAAT	CGAGAGTCT	ATTCAG	ACTTAC	ACTGGATG	CGGCCAA	CAAAATTC	CAACCTTTAGA	240
QY	241	TTTTGTA	CTACTCT	CTACTCT	TATTCCTT	TGGTTGG	GCAACTT	CCAAATAG	GCTCATGTT	300
Db	241	TTTTGTA	CTACTCT	CTACTCT	TATTCCTT	TGGTTGG	GCAACTT	CCAAATAG	GCTCATGTT	300
QY	301	AATCAAT	GATTAGT	GATTATTC	CAGAAATAT	CTTGTTG	TGTTGAC	ATTATTAAT	TATGTCG	360
Db	301	AATCAAT	GATTAGT	GATTATTC	CAGAAATAT	CTTGTTG	TGTTGAC	ATTATTAAT	TATGTCG	360
QY	361	GGGTGAG	ACGGATTAA	ATATCAT	CATGAG	AGCTTAT	CTTTCAT	CTCTCTT	GATTTTGG	420
Db	361	GGGTGAG	ACGGATTAA	ATATCAT	CATGAG	AGCTTAT	CTTTCAT	CTCTCTT	GATTTTGG	420
QY	421	TTTCAG	ATCACTTT	CAGTGT	CACAAG	ATTTTCT	CAGTTTGG	TGTCAT	GAATTTTG	480
Db	421	TTTCAG	ATCACTTT	CAGTGT	CACAAG	ATTTTCT	CAGTTTGG	TGTCAT	GAATTTTG	480
QY	481	AAGTGAG	GTTCCTTAA	ATTTTCA	TATATG	CTTCTT	CTTTTCT	PAGACTAG	CACATGCA	540
Db	481	AAGTGAG	GTTCCTTAA	ATTTTCA	TATATG	CTTCTT	CTTTTCT	PAGACTAG	CACATGCA	540
QY	541	CTTTTCA	CTTTGGGTT	CAAAATG	ACTCACA	AGAAACAA	ATTCAC	TTTTGGGTT	TCACA	600
Db	541	CTTTTCA	CTTTGGGTT	CAAAATG	ACTCACA	AGAAACAA	ATTCAC	TTTTGGGTT	TCACA	600
QY	601	AATTCT	CTTCAG	ATGPACTTT	TTCAC	TGTGA	CTGTCA	GTATAG	GAACAAGGAAT	660
Db	601	AATTCT	CTTCAG	ATGPACTTT	TTCAC	TGTGA	CTGTCA	GTATAG	GAACAAGGAAT	660
QY	661	CAGTTTTT	TAAGGA	CAATGT	ACAGATTT	CACTTTC	CAGAACTCT	TTCTGGT	TGCTCAGTTT	720
Db	661	CAGTTTTT	TAAGGA	CAATGT	ACAGATTT	CACTTTC	CAGAACTCT	TTCTGGT	TGCTCAGTTT	720
QY	721	CAGACT	TTTTTGT	TACCA	AGCTGAT	GCATCA	ATACTT	GTGTTT	CCAAAGCT	780
Db	721	CAGACT	TTTTTGT	TACCA	AGCTGAT	GCATCA	ATACTT	GTGTTT	CCAAAGCT	780
QY	781	ACTGGC	AACTCCT	TAAATG	ATAA	AGAAATA	AAAAAT	CAGATAT	CTCATTTTC	840
Db	781	ACTGGC	AACTCCT	TAAATG	ATAA	AGAAATA	AAAAAT	CAGATAT	CTCATTTTC	840
QY	841	TTGGTT	GGCAGAT	CACAAA	AGGCAACA	AAAGGCT	TAAGCT	CTCTACT	CTTCGGGAGTTA	900
Db	841	TTGGTT	GGCAGAT	CACAAA	AGGCAACA	AAAGGCT	TAAGCT	CTCTACT	CTTCGGGAGTTA	900
QY	901	GGTCAG	GGAC	CCATAT	GAAATG	AAAAAAT	CTTAAAT	TGGGGT	CTCACCAAGAT	960

Simple repeats are identified by repeatmasker (Arian Smit,  
http://ftp.genome.washington.edu/RM/RepeatMasker.html).

## FEATURES

## source

1. .107214

/organism="Arabidopsis thaliana"

/mol\_type="genomic DNA"

/cultivar="Columbia"

/db\_xref="taxon:3702"

/chromosome="I"

/map="mi353"

/clone="F10D13"

/complement(877..1232)

/gene="F10D13.1"

/complement(<877..1232)

/gene="F10D13.1"

/complement(<877..1232)

/gene="F10D13.1"

/codon\_start=1

/product="hypothetical protein, 3' partial"

/protein\_id="AAG60090.1"

/db\_xref="GI:1259778"

/translation="MNPDNPDLSDNDSSAWRELTLTAQSDPFDRTSNILSDFGWNL

HSSDHPHLRFDSDLTQTGVKPTTWTSCSSAAVSVAVTSTNNPSATSSSSBDP

ANSTASAEKTPPPPTP"

1247..1303

/rpt\_family="(GA)n"

1331..1355

/rpt\_family="(GAAA)n"

complement(2071..2094)

/rpt\_family="AT\_rich"

5248..7059

/gene="F10D13.2"

/notes="predicted by genemark hm"

join(<5248..5516,6524..6622,6699..6780,6874..>7059)

/gene="F10D13.2"

join(5248..5516,6524..6622,6699..6780,6874..7059)

/gene="F10D13.2"

/codon\_start=1

/product="hypothetical protein"

/protein\_id="AAG60092.1"

/db\_xref="GI:12597780"

/translation="MKTNRNPINILIVFLLTTARAATNWNTRTRTPVKVOHAYV

AYHRCESPSRYPASMCLELRHSRQPLFSPPPPEIDQSVMNESTVORAKP

EHAAVVDQWETEIEIKRKKYKIEKIVSWEDKKRKAARLKHRTSVEYTKLA

TQRFDERIEIIVASARAHAYESRIKEELKYKAKANLMTTGRKEPSTCL"

complement(7461..8273)

/gene="F10D13.3"

complement(<7461..8273)

/gene="F10D13.3"

complement(7461..8273)

/gene="F10D13.3"

/codon\_start=1

/product="unknown protein"

/protein\_id="AAG60094.1"

/db\_xref="GI:12597782"

/translation="MWFAGSNVLSGMLKSPKDSQAQSCQSDDEVSNISRDEGLE

CPCWESFNIVENPVYVGLHCLQCNCVFLQSAVLRLSSQDRIPIFFVSCPCWQLL

SFRIVYKNGIKFRKRFLLMVMESLNGDRTSHGSLVDNQQSAPTRCPSM

LGNHSSNNLVAPRLRNQSTDLPHDHGNQPSQLFSEKSLDFISFTSKPFYVI

IFLLVFPATPGSLIILALYFLLTILFAPVAGLVLYFAYFILRLVNEITSS"

9624..12967

/gene="F10D13.28"

join(<9624..9944,10045..10122,10256..10311,10628..10747,

10873..10943,11031..11124,11351..11554,11637..11888,

12006..12114,12198..12265,12374..12443,12524..12598,

12713..12804,12892..12967)

/gene="F10D13.28"

join(9624..9944,10045..10122,10256..10311,10628..10747,

10873..10943,11031..11124,11351..11554,11637..11888,

12006..12114,12198..12265,12374..12443,12524..12598,

12713..12804,12892..12967)

/gene="F10D13.28"

/codon\_start=1

/product="unknown protein"

/protein\_id="AAG60116.1"

/db\_xref="GI:12597804"

/translation="MYQITPATPTTGGTPTESGYVTVLDOI PRWSDVEQSSLEDE

TDPEHSPRYANPLASSSEAGSGNMGVSKFPVDHEINSRIYLRGEPWNLEDAVV

NTNENLEASPSGLHVAAGPGLAEQACATGGCTGMAKVTNAYDLFARRVHTVGP

KYAVKYHTAAENALSHCYRSCLELIDSGLOSIALGCIYTEAKNYPRPAAHVAITV

RRFLKQDKISAVVCTTTSDDTEIKRLPLXPRDEHEREVAI SKLPADVGDENG

ETVIDERKIRIQALNKPPTSPPTPLERPSTDLTLRNSHLSDYLDPAFMSLLKD

PDERKQEWKTAQAQSGNFVKLIGLFGDGLGPPPLSAAAEYSLHRSYLAKANSINLSE

IAEMKIVRGKVTGTEHFVMVVAHFLRLCDLDERFLVYVIKEPFPVIQKPYYSI

HAASLQVQDGLGMKRLQILGRKHQNLQAIYVHPTFLKATILMTQFVDNVVV

KKVYADRLQLFKVVPREQLTIPDFVFGHDLVANGKGLIVDPRTKVVYQRP"

13205..15568

/gene="F10D13.4"

/note="contains Pfam profile: PF01535: PPR repeat"

<13205..15568

/gene="F10D13.4"

13205..15568

/gene="F10D13.4"

/codon\_start=1

/product="PPR-repeat protein, putative"

/protein\_id="AAG60097.1"

/db\_xref="GI:12597785"

/translation="MTQYMLPFRSCSLKLVSLQHAHLVLTGRLRDLPLVTKLIESY

AFMGSPDSRIVPEAFVPPVDPFMYGVLIKCNVWCHLLDAAIDLHRLVSETTQISKVF

FFSVLRACAGSREHL SVGKGVHGRILKGVDDDAVLETSLCWMYGTGNLSDAEKVF

GMFVRLVAVSTSLSCLENGEVVRAKRMFKMVDGVPDVTMLSVVSGCAELGCL

RIARSVHQITRKMFDLDTLCLNLTMYSCGDLSSERIPEKIAKNAWSVATKIS

SYNRGEFSEKALRSFSEMIKGIENLVLYSVLSSCGLIGLIRGKSVHGFVRLRL

DPNYSLSLALVELVAECGLKTLVWSDRNIVAWNSLISLYAHRGWVITQALGL

FFQWYTORIKPDFAFTLASSISACENAGLVPLGQIHGVIRTDVSDFFQNSLIDMYS

KSGVDSASTVFNQIKHRSVVTWNSMLCFSGSQNSVEAISLFDYVHSHYLEMVEYTF

LAVIQACSSIGLEKRGVHKKLIISGKDLFTALIDYAKCGDLNAAEVFRFAMS

SRSIVGSMINAYMGHGRISAITFNQWBSGTKNFVNVKGNLVAAGHSGVDEEG

KYFNLKMSFGVSPNSEPACFDLILSRKDEKAYRTIKEMPFADASVWGLSVNGC

RTHQMDI IKAJNDLSIDIVTDDTGYVTLTSLNIYAEGEWEEFRILRSAMKSNLKKV

PGYSATEIDQVFRFGAGEENRIQTDEIVRFLGNLQNLNEEHVVDSD"

complement(15576..18785)

/gene="F10D13.5"

complement(join(<15576..17755,17869..18424,18522..>18785))

/gene="F10D13.5"

complement(join(15843..17755,17869..18424,18522..18785))

/gene="F10D13.5"

/codon\_start=1

/product="unknown protein"

/protein\_id="AAG60100.1"

/db\_xref="GI:12597788"

/translation="MANRALNCVFGSGMGETVHCGSFLSSMRDLSEDISNTCSYMY

CGGDKTLPGYQYNGFSARPPTDSYERDFLKTMLHEAVFKQVYELHRLYRTQKSL

MAEVKGNFVDHLNNEPTPGSGIKGFLFGNSICGEGSTSDCNVGNKVKVLEVPV

KVRTMIDQLPADEYLTHTGNTTCTPPYEQSKVEGENIFPFSHRNDSGSSLLMKN

NGFTDLNEPQCQDSVFSRSSRSLYGLANISHVQGVWEKNTSQNGWVLEAGNG

KSTPRDKLCLPSHVSQVLSAFQPLGYPTSDHLSKLSGRASPKCRQRNPVSXDS

YVSSVASVNPVLNHYGRPESVPSWHSWISWENRSSSVQKPLQANPFLFTNTQV

RADSSAEMSRSNGNLQGSFSSBSAFNFSVNFNLNNGPKGATNGSLCESVMH

QSLKNGKQKCCSSGLPWIKPKPLKNGKTINGLGLDNLASAHQPMEDMDGSSNY

VHPQGRSVTCSNDLNLHCEASVSGVGVKEENKAATHRDLINFCASDEDSG

FSNPVETKATTLINVEAPLTLESBEGGKPEKDEAGDSVDEILIAAEAEIVTIS

LSYCHRTDEAASSSDAVDKPLSWFNTIASCGNDLESKIDACLEARDCECREEC

SGDFEEMATNLNTQTEEDYMPKPLIPEYLKPDGTGMGITSNRPRGRARRGP

KRDFQDILPLGLASLSRLVETDLMFGGIMKATGYWNMSGMARRSRGRKRLVSN

DRAPVCSIAQPMNNSVQVGLDRLSLTGWGNATRRHRRCPCAGTPTPTVLLT"

19062..19085

/rpt\_family="(A)n"

19115..19137

/rpt\_family="(GAGAA)n"

complement(19511..19569)

/rpt\_family="(GA)n"

complement(19621..19663)

9.64; Score 375; DB 8; Length 107214;

repeat\_region

repeat\_region

repeat\_region

repeat\_region

Query Match



ed (03-MAR-2003) The Instit



```

CDS
    /gene="OSJNB0005J14.7"
    <20530..20712
    /gene="OSJNB0005J14.7"
    /codon_start=1
    /product="RNA polymerase beta' subunit-2 5'-partial"
    /protein_id="AA12475.1"
    /db_xref="GI:20146739"
    /translation="QKRVHYFQNKLYFEIQKKLFASEMRDILFLHTELYSSDSV
    TNFYETSESPFPFT"
    repeat_region complement(20788..20809)
    /rpt_family="AT_rich"
    gene 20584..21694

Query Match 4.9%; Score 189.2; DB 8; Length 116932;
Best Local Similarity 49.8%; Pred. No. 6.3e-19;
Matches 694; Conservative 0; Mismatches 613; Indels 87; Gaps 5;

QY 1523 ACATGGACGTGCTCCGCGTACGGGATCTTCAACGCCGACGCGAGCTGTGGAGGAAGC 1582
Db      |||||
QY 91286 ACATGACGACCTGTTCGGGGATGGGATCTTCGCGTCGACGGCGACAAGTGAAGCAGC 91345
Db      |||||
QY 1583 AGAGGAAGACGGCGAGTTTCGAGTTCCCTCCAGAACCTGAGGGATTTTCAGGCCCATG 1642
Db      |||||
QY 91346 AGAGGAAGATCGCCAGCTACGACTTACACCAAGGCGCCCTCCGCGACTTCAGCGCGCAGC 91405
Db      |||||
QY 1643 TGTTCAGAGAGTACTCCCTGAAGCTGCGGGTATCTAGCCAGGCGATCCAGGCGAGCA 1702
Db      |||||
QY 91406 TCTTCAAGAGGAACGGCGCAAGCTCGCGCGCTGCTCCAGCCAGCGCGCTCAAAAC 91465
Db      |||||
QY 1703 AAGTTGGGACATGCAAGT-----GAGATCACTGCTCCCTTGCCATTGCCAAC 1750
Db      |||||
QY 91466 AATCCATGGACTCCAGGTAATCAACCGATCGTAATATATGATGGCGCCAGTGGTGT 91525
Db      |||||
QY 1751 ATGAGCATTTCAACCTGAGACAGAGAGTACCTTCCCGATTCAGGAACCTTTACATGAGG 1810
Db      |||||
QY 91526 GTGTTTGTATCGCATTCAGGAGCTTTGTCGGTGTGTTCTTCAGGGTTTCTTGAAGAGA 91585
Db      |||||
QY 1811 ATGACGTGACCTCATCTGCAAGTTGGTTGGGGTCGAGATCGGACGCTGTCGCG 1870
Db      |||||
QY 91586 GCAACGATGACATCCATCTTCAACATCGCTTCGGCCAGACCTCAACACGCTGGAGCGC 91645
Db      |||||
QY 1871 GATCTCCCGAGAACAGCTTCGGCGAGCGTTCGATCGCGCCAAACATCATCTGCACGCTG 1930
Db      |||||
QY 91646 TCGGCGAGGGCGCCGCTTCGCGCGCGGTTCGACGACGCGAGTTCACCATGCTC 91705
Db      |||||
QY 1931 CGGTTTCAACGCGCTGTGCGCATCAAGAGGTTTCTTCCACGTCGCGGTTCAGAGGCCCTC 1990
Db      |||||
QY 91706 CGCTACCTCAACCGCTTCTGGAAGCTGTGAGGCTCTCAACGTCGCGCGCGAGCGATG 91765
Db      |||||
QY 1991 CTAGCGCAGAGCATCAAGCTCTGGAGAGTTCACCTACAGCGTGTATCCCGCGGAGAG 2050
Db      |||||
QY 91766 CTCAAGAGAGGATCAAGGTTCGTCGCGGGTTCGTGTACAGCTCATCTCCGTGACAGTCC 91825
Db      |||||
QY 2051 GCCAGATCTGTCAGGCGCGGCGCGAGCAAGAGGATGACGTCGACATGACTGT 2110
Db      |||||
QY 91826 GACGAGCTCTCCACACCAAGGC-----ACACGACACTGTAACACTACCTGATAT 91876
Db      |||||
QY 2111 TTGCAATTTTCAGTTCATCGTTCTTGGCGGGATGGACCTGATCTGATGATATATATC 2170
Db      |||||
QY 91877 GTCCAAATTGTCATCCATTTTGTGAGTGATCAACACACTTGAATTAAGGTGAACCTTTTG 91936
Db      |||||
QY 2171 CGTGTGACTGTGAGGACAAATTAATGGCAGATCAAGCAGACATCTCTGTACCGTT 2230
Db      |||||
QY 91937 TGTG-----CAATAGGATTCGAGGCGAGGATATCTCGAAGATT 91975
Db      |||||
QY 2231 CATGAGCTAGGCGGCGGACGACGCGCGCGGTTCGGGACGACAGAGCTCCG 2290
Db      |||||
QY 91976 CAT-----CAAGGACGACTAGGATCTGGGACGTTGATTACAGTACCTGAG 92026
Db      |||||
QY 2291 GGACGTGGTGTCAACTTCGTGATTCGCGGGCGGGAACGACGCGGACGAGCTGTGCTG 2350
Db      |||||
QY 92027 AGACATCATATTGAACATTGTTCATAGCGCGGCAAGGACACACACGCGGCTGCTGCTG 92086
Db      |||||

```

```

QY 2351 GTTACGACATGGCCATGTCTCCACCCGACCTGGCCGAGAGCTGGCGCGGAGCTGTG 2410
Db      |||||
QY 92087 GTTCTGTACATGATGTGCAACACCCGGAAGTACAGGAGAAGTCTGCCAGAACCAT 92146
Db      |||||
QY 2411 CGCGTTTCGAGGCGGACGCGCGCGAGGAGGCGGTTCGCGTCTGTCGTCGCGCGGCGC 2470
Db      |||||
QY 92147 GGAGGCCACCAACGCGCGGCGGCGCTTCCATCGACGAGTTCTGCGAGAGCTGACCGA 92206
Db      |||||
QY 2471 TGACGCGCAGCAGCAAGGCGTTTCGCGCGCGCGGTGGCGGAGTTTCGCGGCGCTCTC 2530
Db      |||||
QY 92207 CGAGGCACTGAACAAGATGACATATCTGCACTGACGCTGACGAGAGACGCTCAGGCT 92266
Db      |||||
QY 2531 CGACAGCTTCGCGCAAGCTGCTTACCTCCACGCTCGCTCAACGAGACGCTCCGCTGTA 2590
Db      |||||
QY 92267 CCTGCACTTCCTGCTGATTAATTAATTAATCAAGTAAATCAACGAATTAATCT 92326
Db      |||||
QY 2591 CCCCCTGCTTCCTGAGTGAAGCGCGCGCGACCGGACCTCCGCTCCAGAGCAGCATG 2650
Db      |||||
QY 92327 CTTGCCCAATGCCMAATGAATCGAGTATAGTAACAACAGTAGTGTC----- 92374
Db      |||||
QY 2651 CAGTGAAGTGAACCTGAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCA 2710
Db      |||||
QY 92375 -----CTTTGTTTCTGACGATGTTTTCAGGATAACAACAGT 92410
Db      |||||
QY 2711 TCCTGAGGACGACGCTGCTCCGCGACGCGCAAGGTGAGGCGCGCGGAGTGGTGA 2770
Db      |||||
QY 92411 GCTTCTCAGACGATGATTGCCCCAAGGATTCAACGTCAGCAAGGGGACATCGTGTCT 92470
Db      |||||
QY 2771 AGTGGCTTCTGATGAGGCGGAGTGAATCAACCTGGGCGCGCGCGGAGGAGTTC 2830
Db      |||||
QY 92471 ACATCCCTTACGCGATGCGCGGATGAGAGGTTGTGGGCGCAAGACGCTGAATCTTCC 92530
Db      |||||
QY 2831 GGCGGAGCGGTGGATCAACGAGGATGCGCGCTTCGCAACGCTCCGCTTCAAGTTCA 2890
Db      |||||
QY 92531 GGCTGAACGTTGGCTCGATGAGAACGCGCTCTTTTCAGGAGGAGAGCCGCTTCAATTT 92590
Db      |||||
QY 2891 CGCGCTTCCAGGCG 2904
Db      |||||
QY 92591 CAGCTTCCAGTGT 92604
Db      |||||

RESULT 11
LOCUS AE017114.1 300029 bp DNA linear PLN 06-JUN-2003
DEFINITION Cryza sativa (japonica cultivar-group) chromosome 10, section 68 of 77 of the complete sequence.
ACCESSION AE017114 AE016959
VERSION AE017114.1 GI:31433165
KEYWORDS
SOURCE Cryza sativa (japonica cultivar-group)
ORGANISM Cryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 300029)
AUTHORS The Rice Chromosome 10 Sequencing Consortium
CONSTRM In-depth view of structure, activity, and evolution of rice chromosome 10
TITLE Science 300, 1566-1569 (2003)
JOURNAL 2 (bases 1 to 300029)
AUTHORS Buell, C.R., Wing, R.A., McCombie, W.R., Messing, J. and Yuan, Q.
TITLE Direct Submission
COMMENT Submitted (05-MAY-2003) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
This is the pseudomolecule for rice chromosome 10, which was constructed by resolving discrepancies between overlapping BACs, trimming the overlap regions, and linking the unique sequences to form a contiguous sequence. Genes in individual BAC clone were identified by a combination of several methods: Gene prediction programs, searches of the complete sequence against a peptide database and EST databases. Genes with similarity to other proteins are named after the database hits. Genes without significant

```



mRNA	/locus_tag="OSJNB0005J14.22"	11271	CGTGTGACTTGTGAGGACAAATTAATAATGGCAGATGAAGCAGACGATCCTCTCGAGGTT	2230
	/note="predicted by fgenesh"	12578	TGTG	12540
CDS	complement(<44575..>45057)	2231	CATCGAGCTAGCGGAGCGCGGCGAGCGAGCGGGCGCGCTTCGGGGACGACGAGAGCTCCG	2290
	/locus_tag="OSJNB0005J14.22"	12539	CAT	12489
gene	/product="hypothetical protein"	2291	GGAGCTGGTCTCACTTCGTGATCGCGGGGGGACGACGAGCGGCGAGCGCTCGTG	2350
	/protein_id="AAP54715.1"	12428	GTTCTGTACATGATGTGCAACACCCCGGAGTACAGGAGAAGATCTGCCACGAGCCAT	12369
tRNA	/db_xref="GI:31433173"	2411	CCGCTTCAGGCGGAGCGCGCGGAGGAGGCGGCTTCGCGCTTCGCTCGCGGGCGGC	2470
	/note="similar to C terminus of glutathione S transferase"	12368	GGAGGCCACCAACGCGCGGAGGCGCTTCATCGACGAGTTCTCGCAGAGCTGACCGA	12309
Query Match	GB:AAF29773 GI:6856103 (Gossypium hirsutum); EST BE530893	2471	TGACGCGGACGACGAGGCGTTGCGCGCGCGCGCTTCGCGCTTCGCGGGCGAGTTCGCGGGCTCTCACCTA	2530
	from this gene	12308	CGAGGCACTGAACAGAGATGCATATCTGCACGCTGCACCTGACGAGAGCGCTCAGGCTATA	12249
Best Local Similarity	/pseud	2531	CGACAGCTCGGCAAGCTGCTCTACCTCCACGCTCGCTCACCGAGAGCGCTCCGCGCTGTA	2590
	<52388..>52459	12248	CCCTGAGTTCCACTGTGAATTAATTAATGATCAAGTAAATCAACGAAATTAATTCCT	12189
Matches 694; Conservative 0; Mismatches 19; Indels 87; Gaps 5;	/locus_tag="CHRI10v09222002.LRNA-Val-3"	2591	CCCCCGCTCCCTCAGGTGAGCGCGCGCGGAGCGGAGCGGCTTCGCTCCGCTCCGAGGACGACGATG	2650
	52388..52459	12188	CTTGCCCAATGCAATGAATCGAAGTATAGTAAACAAAGTAGTGTCA	12141
1523	/locus_tag="CHRI10v09222002.LRNA-Val-3"	2651	CAGTGAGTGAGACTGATGCAATGCAATGACATGCACTTCGCGCGCGGAGGAGGAGGAGG	2710
		12140	-----CTTTGTTTCTGACGATGTTTTCAGGATAACAAGCACT	12105
13229		2711	TCCTGGAGGACGCTGCTGCGGAGCGGAGCGAAGGTGAGGCGCGGAGGATGTTGACCT	2770
		12104	GTTCTCAGACGATGATTTGCCCCACGATTCACGTCAGCAAGGGGACATCGTGTCT	12045
1583		2771	AGTGCCCTTACTGATGGGCGGATGGAGTACAACTGGGGCCCCCGAGCGGAGCTTCC	2830
		12044	ACATCCCTACGCGATGCGCGGATGGAGAGCTTGTGGGGCAAGACGCTGAATCTTCC	11985
13169		2831	GGCGGAGCGGTGATCAACGAGGATGGCGCTTCGCAACGCGTCCGCTCAAGTTCA	2890
		11984	GGCTGAACGTTGGCTCGATGAGACGCGCTCTTTACGAGGAGAGCGCTTCAAAATTTA	11925
1643		2891	CGGCGTTCCAGGCG 2904	
		11924	CAGCTTCCAGGTG 11911	
13109		RESULT 12		
		AK118803		
1703		LOCUS		
		AK118803		
13049		DEFINITION	Arabidopsis thaliana At1g69500 mRNA for unknown protein, complete cds, clone: RAPL21-14-E14.	
		AK118803		
1751		ACCESSION	AK118803	
		VERSION	AK118803.1	
12989		KEYWORDS	FLI CDNA; CAP trapper.	
		SOURCE	Arabidopsis thaliana (thale cress)	
1811		ORGANISM	Arabidopsis thaliana	
			Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.	
12929		REFERENCE	1	
		AUTHORS	Seki, M., Iida, K., Satou, M., Sakurai, T., Akiyama, K., Ishida, J., Nakajima, M., Enju, A., Kamiya, A., Narusaka, M., Carninci, P., Kawai, J., Hayashizaki, Y., and Shinozaki, K.	
1703		TITLE	Arabidopsis thaliana full-length cDNA	
		JOURNAL	Published Only in Database (2002)	
12638		REFERENCE	2 (bases 1 to 1695)	

**AUTHORS** Seki, M., Iida, K., Satou, M., Sakurai, T., Akiyama, K., Ishida, J., Nakajima, M., Enju, A., Kamiya, A., Narusaka, M., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.

**TITLE** Direct Submission

**JOURNAL** Submitted (25-NOV-2002) Motoaki Seki, RIKEN Genomic Sciences Center; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: msek@gs.riken.go.jp, URL: http://pfweb.gsc.riken.go.jp, Tel: 81-45-503-9625, Fax: 81-45-503-9586)

**COMMENT** An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al. (1998) Plant J. 15:707-720; Seki et al. (2002) Science 296:141-145). cDNA cleaved with BamHI and XhoI was ligated to modified Lambda FLC-1-E vector (Carninci et al. (2001) Genomics 77:79-90) digested with BamHI and SalI. This clone is in a modified pluescript vector. Please visit our web site (http://pfweb.gsc.riken.go.jp/) for further details.

**FEATURES** Location/Qualifiers

source 1..1695

organism="Arabidopsis thaliana"

mol\_type="mRNA"

cultivar="Columbia"

db\_xref="taxon:3702"

chromosome="1"

clone="RAF121-14-E14"

note="common name: Thale cress"

1..1695

gene /gene="Atlg69500/F10D13\_15"

CDS 110..1546

gene /gene="Atlg69500/F10D13\_15"

codon\_start=1

product="unknown protein"

protein\_id="BAC43393.1"

db\_xref="GI:26452620"

translation="MHDWLVEVLYNSRTVVVMPPTTYIADPNINVEYLVKTFNSY PKGEYHSYVMDLGDGIFNSDGLRQKRTASPEFASKNLROPTVVPKYSIKLF TILSOASPEQVMDQELLMRWLDSICKVGFVEIGTLAPLPENHFKAFTANII VTLRFDPMKKFENLGEALLGSKIKVNDYFYSVIRKAELELBAQISPTNNN NNNKVKHDIILSRFEISDDPDSKTERSLDRDIVNFVIAGRTDTATITLWIMM NENAEKLYSELLEKESAEATNSLHQYDTEFNSEKVTPEFAGLLNYSGLKH YLHAVITETRLYPVAPQDPKGLVDDMLPNGTKVAGMVTYVPYSGRMEYNGSD AALFKPERWLKGVFNASPKPTFAQAPRICLGDKSAYLQMKMAMAILCRFYKFL VPHVFKYRMMLTSLNAHLKLVTSRS"

**ORIGIN**

Query Match 4.8%; Score 186; DB 8; Length 1695;

Best Local Similarity 50.7%; Pred. No. 2.6e-18;

Matches 959; Conservative 0; Mismatches 570; Indels 361; Gaps 7;

QY 1182 TCATGGATCCTGGTCCAGAGGTGGAGCCCTGAGAGACGAGAAAGCCCGAGATCATGGCCA 1241

DB 8 TCATGGATCCTTTTACACCGATGGGACAAAGAAACAGAGTGGTCCAAAGACTTGGCCT 67

QY 1242 GTCATGGTGGCAACGGTGGAGCAGCTGAGGAAGTACACCGGATGACGACCTGGCTGTGC 1301

DB 68 TTGGTTGAGACAGCGATCGAAAGTTAACTAATTTGATCGAATGATGATGGCTCGTT 127

QY 1302 GGGTACCTGTACGGCACAGGACAGTACCGTGGACATGCGGTTCACTTCTTACACCTAC 1361

DB 128 GAGTATCTTTATAAATCAAGAACAGTAGTGGTTCCTCAATGCCATTCCTACTTATACATAC 187

QY 1362 ATCGTGCACCGGTGAATTCGAGCATGTCCTCAGACATACTTCCACCAATTACCCCAAG 1421

DB 188 ATAGCTGATCCCAATCAACGTTGAATATGCTTCGAAACAAACACTTCCCAATTACCCCAAG 247

QY 1422 GTAAATGACCTGAACCTCACTGATGTTTCAGTCTTCGGAAATCAGAGCTGAAAGCTGAATCG 1481

DB 248 GAGAGACT----- 256

QY 1482 AATGTCCTGAAACACCGTGTAGGGAATCGGTGACATCTTACATGAGCGTGTCTCTCGG 1541

DB 257 -----TATCATTTCTTATGGAAGTATTGTTGGG 285

QY 1542 TGACGGCACTTTCAACGCCACGCGAGCTGTGGAGGACGAGGAGGAGCGGCGAGTTT 1601

DB 286 AGATGGGATTTTCAATTTCTGATGGAGAGCTTTGGAGGAAACAGAGAAAACCGCGAGTTT 345

QY 1602 CGAGTTTCGCTTCAAGAAACCTGAGGGATTTTCAGCGCAATTTGTGTTTCAGAGAGTACTCCCT 1661

DB 346 CGAATTTGCTTCCAGAAATCTTAGAGATTTTAGTACTGTAGTGTATTAAGAGATAGTCT 405

QY 1662 GAAGCTCTGGGTATATCTGAGCCAGGATCGAAGGAGGCAAGTGTGTGACATGACAGT 1721

DB 406 TAAACTCTTCACTATCTCTTTCCCAAGCATCTTTCAAGAGCAACAGATAGATATG----- 460

QY 1722 GAGATCACTGCTCCCTTGCCATTGCCAAACATGASCATTTCAACTGAGACACGAGAGCTA 1781

DB 461 ----- 460

QY 1782 CTTTCCCGATTCAGAACTTTTACATGAGATGAGCTGGACTCCATCTGCAAGTTGGGT 1841

DB 461 -----CAGGAATTGTTGATGAGATGACTTTGGATTCAATATGTAAGTGGGAT 509

QY 1842 TCGGGTTCGAGATCGGCAACGCTGTCGCCGATCTCCCGGAGACAGCTTCGGCAGGCGT 1901

DB 510 TTGGTGTGAGATAGGAGCATTTGGCTCCAGAGTTTACCAGAGAAATCACTTTGCTTAAGGCTT 569

QY 1902 TCGATCGCGCAACATCATCTGCTACGCTGGGTTTCATCGACCCCTGTGGCGCATCAAGA 1961

DB 570 TTGATACCGCAATATATCGTAACATTCGTTTCATAGATCTCTTTTGGAGATGAAA 629

QY 1962 GTTTCCTTCCAGCTCGGTTCAGAGCCCTCTCTAGCGCAGAGCATCAAGCTCGTGGACGAGT 2021

DB 630 AGTTCTTAAACATAGGCTCTGAGGCATTTGCTTGTAAAGCATAAAGTAGTCAATGATT 689

QY 2022 TCACCTACAGGCTGATCGCGGAGGAGCCGAGATCGTTCGAGGCGCGGCGGCGGCGCA 2081

DB 690 TCACCTATTCAGTCATAAGAAAGGAGGAGAGAGTATTAGAGGCAC----- 737

QY 2082 AACAGGAGAGGTACGTGCATGACTGTTTCGATTCTTCAGTTCATCGTCTTGGCGCGG 2141

DB 738 ----- 737

QY 2142 ATGGACCTGATCCTGATTGATTATATATTCGCTGTGACTTGTGAGGACAAATTAATAATGG 2201

DB 738 -----AAATATCTCCTACCAACAACAACAACAACAACAACAACAACA 775

QY 2202 CAGATGAAGACGACATCCTGTCAGGTTTCATCGAGCTAGCGAGGCGCGGAGCACGCGC 2261

DB 776 AAGTTGAAGCAGATATATCTCTGAGGTTTATCGAGAT-----CAGTGACATCCA 826

QY 2262 GCGGCTTTCGGGACGACAGAGAGCTCCGGGACGCTGCTCAACTTCGATCGCCGCGG 2321

DB 827 GATAGCAAGAAACAGAGAAAGCCTAAGAGATATTTGCTCAACTTTGTTAATAGCTGGA 886

QY 2322 CGGACACGACGGGACGACGCTGTCTGTGGTTACGACATGGCCATGTCACCCGAC 2381

DB 887 AGAGATACAACAGCAACAACCTCTCACTTGGGCTATATACATGATAATGAAATGAATAAT 946

QY 2382 GTGSCCGGAGAGCTGCGCGGAGCTGTGCGGTTTCGAGGCGGAGCGCGCGGAGGAG 2441

DB 947 GTTGTGAGAACTATATCTCAGAGCTACAAGAACTCGAAAGAAAGCGGAGGAGCT 1006

QY 2442 GCGCTGCGGCTGTGTCCTTGGCGGCGCTGACGCGGACGACGAAGCGGTTTCGCGCCCGC 2501

DB 1007 AATACATCGTTGCATCAATA-----CGATACAGAGGATTTCAATTCCTTCAACAGAGAAG 1060

QY 2502 GTGGCGAGTTGCGGGGCTCTCTCACCTACGACAGCTCGGCAAGCTGGTCTACTCTCAC 2561

DB 1061 GTAAACAGAAATTCGAGGACTATTGAACCTACGATTCTTTAGGAAACATTCCTACTTACAT 1120

QY 2562 GCTCGCTCACGAGACGCTCGCGCTGTACCGCGCTCCCTCAGTGCAGGCGCGCCCGAC 2621

DB 1121 GCTGTGATCACAGAAACACTTTCGTCTTACCTCGCTTCCT----- 1162

QY 2622 ACGGACCTCCGGTCCAGAGCAGCATGTCAGTGTGAGTGCACCTGAATGCAATGCATGC 2681

Db 1163 ----- 1162

QY 2682 ACTTGGCGCGCGAGACCCCAAGGGGATCTCTGGAGGACGAGCTGCTGCGGACGGGAC 2741

Db 1163 -----CAGATCCGGAAGAGATATTAGAGATGATATTACCAATCGGAC 1209

QY 2742 GAAGGTGAGCGCGCGGAGTGTGACGTACGTGCGCTACTCGATGGCGGAGTGAAGTA 2801

Db 1210 AAAAGTAAAGCTGGAGGATGTCACATATGATACCTTACTCAATGGTGGTATGAATA 1269

QY 2802 CAACCTGGGCGCGAGCGCGGAGCTCCGCGCGGAGCGGTGATCAACAGGATCGCGC 2861

Db 1270 CAATTGGGGATCAGATGCGACATGTTTAAACCGGAAGATGGCTTA---AAGATGGAGT 1326

QY 2862 GTTCCGCAACCGCTCGCGGTTCAAGTTTCAGCGGTTCCAGCGCGGCGGAGGATCTGCGCT 2921

Db 1327 CTTTCAAAATGATCACCATTCAAGTTTCAGCTTTCAGCGTGGACCAAGGATATGTTT 1386

QY 2922 GGCAAGGACTCGCGTACTCGAGATGAAGATGGCGCTGGCGATCCTCTTGGCGTTCTA 2981

Db 1387 GGCAAGGATTCAGCTTATCTGCAAAATGAAGATGGCAATGGCAATCTTTGCGAGATTTCA 1446

QY 2982 CAGCTTCGCGCTCTGGAGGGGACCGGTCAGTACCGCATGATGACCATCTCTCCAT 3041

Db 1447 TAAGTTTCATTGGTGCAAAATCATCTGTCAAAGTATCGGATGATGACGATCTTATCCAT 1506

QY 3042 GGCGCAGCGGCTCAAGTCCGCGTCTCTAG 3071

Db 1507 GGTCTATGGCTTGAAGTTACTGTATCCAG 1536

RESULT 13

AC078894/c 140681 bp DNA linear PLN 20-DEC-2002

LOCUS Oryza sativa chromosome 10 BAC OSJNBa0096G08 genomic sequence,

DEFINITION complete sequence.

ACCESSION AC078894

VERSION AC078894.11 GI:273111268

KEYWORDS HTG.

SOURCE Oryza sativa (japonica cultivar-group)

ORGANISM Oryza sativa (japonica cultivar-group)

REFERENCE 1 (bases 1 to 140681)

AUTHORS Buell, C.R., Yuan, Q., Ouyang, S., Liu, J., Gansberger, K., Kim, M.M., Overton, II, L.L., Bera, J.J., Tsitrin, T., Krol, M.I., Jarrahi, B.B., Jin, S.S., Koo, H., Zismann, V., Heiao, J., Blunt, S., Vanaken, S.S., Utterback, T.T., Feldblyum, T.V., Yang, Q.Q., Haas, B.J., Suh, B.B., Peterson, J.J., Quackenbush, J., White, O., Salzberg, S.L. and Fraser, C.M.

TITLE Oryza sativa chromosome 10 BAC OSJNBa0096G08 genomic sequence

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 140681)

AUTHORS Buell, R.

TITLE Direct Submission

JOURNAL Submitted (08-AUG-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

REFERENCE 3 (bases 1 to 140681)

AUTHORS Buell, R.

TITLE Direct Submission

JOURNAL Submitted (15-NOV-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

REFERENCE 4 (bases 1 to 140681)

AUTHORS Buell, R.

TITLE Direct Submission

JOURNAL Submitted (25-APR-2002) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, rbuellt@tigr.org

REFERENCE 5 (bases 1 to 140681)

AUTHORS Buell, R.

TITLE Direct Submission

JOURNAL Submitted (18-DEC-2002) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

REFERENCE 6 (bases 1 to 140681)

AUTHORS Buell, R.

TITLE Direct Submission

JOURNAL Submitted (20-DEC-2002) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, rbuellt@tigr.org

On Dec 20, 2002 this sequence version replaced gi:27228824.

Address all correspondence to:rice@tigr.org

BAC clone OSJNBa0096G08 is from Oryza sativa chromosome 10

The orientation of the sequence is from SP6 to T7 end of the BAC clone.

Genes were identified by a combination of several methods: Gene prediction programs including Fgenesh (<http://www.softberry.com/>), genscan and Genscan+ (Chris Burge, <http://CCR-081.mit.edu/GENSCAN.html>), GeneMarkMM (Mark Borodovsky, <http://genemark.biology.gatech.edu/GeneMark/>), and GeneSplicer (Mihaela Pertea and Steven Salzberg, [contact mpertea@tigr.org](http://contact.mpertea@tigr.org)), searches of the complete sequence against a peptide database and the plant EST database at TIGR (<http://www.tigr.org/tdb/tgi.shtml>). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by repeatmasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>).

Location/Qualifiers

1. 140681

/organism="Oryza sativa (japonica cultivar-group)"

/mol\_type="genomic DNA"

/cultivar="Nipponbare"

/db\_xref="taxon:39947"

/chromosome="10"

/map="E1363S"

/clone="OSJNBa0096G08"

/notes="japonica cultivar-group"

585. 1903

/gene="OSJNBa0096G08.1"

/notes="predicted by fgenesh"

join(<585. 599,1096. 1228,1707. >1903)

/gene="OSJNBa0096G08.1"

join(585. 599,1096. 1228,1707. 1903)

/gene="OSJNBa0096G08.1"

/codon\_start=1

/product="hypothetical protein"

/protein\_id="AA00895.1"

/db\_xref="GI:273111269"

/translation="MKYRKLICRGWDPDPEDISKYAMIEWFKSQMQEAGIVDREDLY

FTFSDTQLGKLLPQGLSVLAQPVLLNPDPNTNSEIKARGFILLASNSYAYSEKRV

WIRTANKFQCT"

3357. 5378

/rpt\_family="(CAG)n"

complement(5695. 5715)

/rpt\_family="AT\_rich"

5882. 5915

/rpt\_family="(CACCT)n"

complement(6798. 6848)

/rpt\_family="AT\_rich"

7098. 9490

/gene="OSJNBa0096G08.2"

/notes="contains Pfam profile PF00514 (Armadillo beta catenin like repeat); EST C71879, AU181935, AU161982, B1806399, AU095513, AU032071 from this gene"

join(7098. 7339,7509. 9490)

/gene="OSJNBa0096G08.2"

join(7330. 7339,7509. 9226)

/gene="OSJNBa0096G08.2"

/codon\_start=1

```
/product="putative armadillo repeat containing protein"
/protein_id="AA00697.1"
/db_xref="GI:27311271"
/translators="MVSDBVRCSSFDGMSLHCCPAAGPPAPAGTABELLERARSL
VPAALDARARATGGRGWKVIARLBRVPPCLDLSHPCFKNLSLCRELLQSVAAVL
AEALGARGCREPRAGKLOMSDLDLALACKLDLNRDCCALLIKTGVLSDATVPPVAP
AANAQDRELLARLQIHAENAKHRVNDGLDALREDEKSVLSALGRGNVAALVQL
LITAPKIREKAATVLCLLAESGCECLLVSEGAALPLILRVSGSLVGREKAVITIQ
RLNSPDIAAGVGVRLPIDICQTDGSISSQSAAGALKNLISAVPEVQALAEQI
VRVNVLDGCVLGCKEYAECLQSITSSNDGLRRVAVSEGGRLSLAYLDGPLQOE
SAVCALNLYSSATSPSLVSLVPLRVHVLREGSVGAQAAAAAICRVSSSENKR
LVGSHGCMPLLVLEKASNGAREVAQAQVASMISCLANARDIKKDEKSVNVLVQLE
PSPQNTAKKVAISCLLTSASRCKKLMISHGALGYLLKLSMDVAGAKKLEKLEKRG
KLRNLFRRK"
```

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

/rpt\_family="Oryza"

complement(11631..11725)

/rpt\_family="AT-rich"

complement(11631..11725)

/rpt\_family="tandem"

complement(11631..11725)

/rpt\_family="Oryza"

complement(11631..11725)

/rpt\_family="Oryza"

complement(11631..11725)

/rpt\_family="Oryza"

complement(11631..11725)

/rpt\_family="tandem"

complement(11631..11725)

/rpt\_family="TA"

complement(11631..11725)

/rpt\_family="Oryza"

complement(11631..11725)

/rpt\_family="Oryza"

complement(11631..11725)

/rpt\_family="Oryza"

complement(11631..11725)

/rpt\_family="Oryza"

complement(11631..11725)

/rpt\_family="Oryza"

complement(11631..11725)

/rpt\_family="Oryza"

complement(11631..11725)

/rpt\_family="Oryza"

complement(11631..11725)

/rpt\_family="Oryza"

complement(11631..11725)

/rpt\_family="Oryza"

complement(11631..11725)

/rpt\_family="Oryza"

complement(11631..11725)

/rpt\_family="Oryza"

complement(11631..11725)

/rpt\_family="Oryza"

complement(11631..11725)

/rpt\_family="Oryza"

complement(11631..11725)

/rpt\_family="Oryza"

complement(11631..11725)

/rpt\_family="Oryza"

complement(11631..11725)

/rpt\_family="Oryza"

complement(11631..11725)

/rpt\_family="Oryza"

complement(11631..11725)

/rpt\_family="Oryza"

complement(11631..11725)

/rpt\_family="Oryza"

Query Match

4.5%; Score 174.4; DB 8; Length 140681;

Best Local Similarity 49.9%; Pred. No. 1.1e-16;

Matches 696; Conservative 0; Mismatches 616; Indels 82; Gaps 7;

1529

138438

1589

138378

1649

138318

1709

138258

1766

138198

1811

138138

1871

138078

1931

138018

1991

137958

2051

137898

2111

137838

2171

137783

2231

137733

2291

137686

2291

137686

2291

137686

2291

137686

2291

137686

2291

137686

2291

137686

2291

137686

2291

137686

2291

137686

2291

137686

2291

137686

137685 AGACATCATCTGAAATGTCATAGCCGCAAGACACACAGCTGGGCGCTGCTTG 137626  
2351 GTTCAGGCATGGCATGCTCCACCCGACGCTGGCGGAGAGCTGCGCGGAGCTGTG 2410  
137625 GTTCTGTACATGGTGTGACACCCGAGAGTCCAGGAGAGATCTGCCAGGACCAT 137566  
2411 CGCTGTGAGGCGGCGCGCGGAGGAGGCGGTGCGCTGCTGTGCGCTGCGGCGGCG 2470  
137565 GGTGGCCACAGCGCGCGGACACCGCTTCCGTCGACGAGTCTTTCAGAGCTTCACTGA 137506  
2471 TGACGCCGACGACAGGCGCTTCCGCGCGCGCTGGCGGAGTTCGGGGCTCTCTCACCTA 2530  
137505 CCAAGCGCTGAACACATGCACTACCTGTCACGCGGCACTGACGAGACGCTCCGGCTTTA 137446  
2531 CGACAGCTCGGCAAGCTGCTTCACTCCACGCGCTGCTGCTGACGAGACGCTCCGCTGTA 2590  
137445 CCTTCAGTTCCTAGTGTGATTAATATGATCAGTTCATCTA----- 137402  
2591 CCGCGCGTCCCTCAGTGAGCGCGCGGACGACGCGACCTCCGCTCCAGACGACGATG 2650  
137401 -----CCATCGAGCTAATTAATCAACGACATTTCTCTCATCGAGTGGCTAACAAA 137351  
2651 CAGTGAGTGCACCTGATGCAATGCACATGCACTTGGCGCGGAGGACCCCAAGGGGA 2710  
137350 CGTCTCTGTGATG-----ATGATCTTTTCTTTTTCAGGAAACAGACAGT 137303  
2711 TCTGAGGACGACGCTGCTGCGGAGCGGACGAGGTGAGGCGCGCGGATGGTCACT 2770  
137302 GCTTTTCAGACGATGATTGCCCCACGGTTTAACTGTCAGCAAGGGGACATGCTTCT 137243  
2771 ACCTGCGCTACTGATGGGCGGATGAGTAACTGCGGCGCGGACGCGGCGAGCTTCC 2830  
137242 TCATCCCTACGCGATGGCGGATGGAGAGCTTGTGGGCAAGACGCTGATATCTTCC 137183  
2831 GCGCGGAGCGGTGATCAACGAGGATGGCGCTTCCGCAACGCTGCGCGCTTCAAGTTCA 2890  
137182 GGCTGACCTTGGCTGCGAGAAACGGCTCTTTCAGCAGGAAGCCGCTTCAAGTTTA 137123  
2891 CGCGCTTCCAGGCG 2904  
137122 CAGCAITTCAGGTG 137109

RESULT 14  
AE017113/c  
LOCUS  
DEFINITION  
Oryza sativa (japonica cultivar-group) DNA linear PLN 06-JUN-2003  
77 of the complete sequence.  
ACCESSION  
AE017113 AE016959  
VERSION  
AE017113.1 GI:31433100  
KEYWORDS  
Oryza sativa (japonica cultivar-group)  
SOURCE  
Oryza sativa (japonica cultivar-group)  
ORGANISM  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzaceae; Oryza.  
REFERENCE  
1 (bases 1 to 300029)  
The Rice Chromosome 10 Sequencing Consortium  
In-depth view of structure, activity, and evolution of rice  
chromosome 10  
JOURNAL  
Science 300, 1566-1569 (2003)  
REFERENCE  
2 (bases 1 to 300029)  
Buell, C.R., Wing, R.A., McCombie, W.R., Messing, J. and Yuan, Q.  
AUTHORS  
Direct Submission  
TITLE  
Submitted (05-MAY-2003) The Institute for Genomic Research, 9712  
JOURNAL  
Medical Center Dr, Rockville, MD 20850, USA  
COMMENT  
This is the pseudomolecule for rice chromosome 10, which was  
constructed by resolving discrepancies between overlapping BACs,  
trimming the overlap regions, and linking the unique sequences to  
form a contiguous sequence. Genes in individual BAC clone were  
identified by a combination of several methods: Gene prediction

programs, searches of the complete sequence against a peptide  
database and EST databases. Genes with similarity to other proteins  
are named after the database hits. Genes without significant  
peptide similarity but with EST similarity are named as unknown  
proteins. Genes without protein or EST similarity, that are  
predicted by more than two gene prediction programs over most of  
their length are annotated as hypothetical proteins. Genes  
encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy,  
http://genome.wustl.edu/eddy/tRNAscan-SE/).

FEATURES  
source

Location/Qualifiers  
1..300029  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="genomic DNA"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/chromosome="10"  
misc\_feature  
c1..97039  
/note="Chromosome Sequence Derivation: nucleotide sequence  
in this region was derived from BAC clone OSJNB0018B10  
(GB:AC051634)."  
complement(<1466..>3459)  
/locus\_tag="OSJNB0018B10.7"  
/note="predicted by genemark hm"  
complement(join(<1466..1579,2443..2466,3325..>3459))  
/locus\_tag="OSJNB0018B10.7"  
complement(join(1466..1579,2443..2466,3325..3459))  
/locus\_tag="OSJNB0018B10.7"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="AAP54654.1"  
/db\_xref="GI:31433101"  
translation="MLAPWYGRHFRHWRPLRPSOHLHDSPIATVARARRQCGRRHV  
GIKEADREKERNAYVDLAGCGFEIEGYPVVDYESALQTAKSTTVR"  
complement(<3713..>5670)  
/locus\_tag="OSJNB0018B10.8"  
/note="predicted by genemark hm"  
complement(join(<3713..3773,4324..4479,4573..4619,  
4691..4789,4866..5151,5266..5474,5560..>5670))  
/locus\_tag="OSJNB0018B10.8"  
complement(join(3713..3773,4324..4479,4573..4619,  
4691..4789,4866..5151,5266..5474,5560..5670))  
/locus\_tag="OSJNB0018B10.8"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="AAP54655.1"  
/db\_xref="GI:31433102"  
translation="MEKTVAYLFESTATTHAHRDVAYKHVNMDCLOQOHTEWLPHY  
TNEASLTUNSCNRDSDYFMYQCPICFWAYVYHLPVRMRQFGKKQWPFVEDISTG  
VELHKVTKKVDGWLGNHRYIDWRTAGNDRYETTHQHLFSEYDLRWLHRTYLF  
LRPTTEADKEDDRSDERRNPVDVTRVGYQMEHAPLRDRSVRELLRVNTEIGHALQ  
APRGEDTENTLRNVLEKVRQCRKLAAALGRGIEVREGVGSSSSRIKQGRGKGA  
APSDDDDEDEDEDEYVADAEIDMEAGSNHFGATVGKGGAS"  
complement(<8248..>9465)  
/locus\_tag="OSJNB0018B10.9"  
/note="predicted by genscan"  
complement(<8248..>9465)  
/locus\_tag="OSJNB0018B10.9"  
complement(<8248..>9465)  
/locus\_tag="OSJNB0018B10.9"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="AAP54656.1"  
/db\_xref="GI:31433103"  
translation="MDWSLIPADLLAVFARLPDRADARFRVACVACGAAASWRPR  
PWFVGRSTRDFRGAAMSSFWLSRRRGGGLVFFAAVAVAGLEFFSSSHGVLVPR  
PMATKAVALVNPVTGRTLRPIPGFKRHWLVDVLSADPTDADAAVAVGFFPN  
CLAYSSAAGAWTPLGFAAGVGHFRGFYVAFKSQLCVCDEATVAVTLEQL  
IDDDGGGNDVTRGVVTHVCEGDELLVSHVNDLERNPDTTTFGGDDHDDGGS  
SSSDSGDGRVVEVAGVAGVAVRVEDLRSLRFLGRNHAFALSPPEPSCRA  
NCVLYDQCGHPVRVFMNADAEWEPEAAVAVVARYARDETTFPPDDGRRDAQS  
AGWARRGWFFPKY"  
complement(<10339..>11589)  
/locus\_tag="OSJNB0018B10.10"  
gene

```

/note="predicted by genemark hmm"
mRNA complement (<10339. .>11589)
CDS /locus tag="OSJNBb0018B10.10"
/locus tag="OSJNBb0018B10.10"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAP54657.1"
/db_xref="GI:31433104"
/translation="MAAEEDWASLEOGLLHDFILRLDDANAAFRFVAVCSQWRAAAAG
AGVVFPRPLAVAPORRHAIIPRRRRRRRRRRRFRKRVDDGVRSLRPRGGSPAPA
AVLSVSRGWAMKGGESHLVRAVSDEVPPLPCGSRYSLESEVPLPCGSRVLRVE
LSKVLDPDPAAGASGEWATAFAMSHHQQLLWPHTRRVAFCRAGDEWELFLKAIQA
QRYGLFFQVQSVRRPSCPLLAGRRRRRLFTSRVYKLEFAADGSPVFGFTK
VESIGELFVSRSHAFPLPASGAASPASRIASTTWSDYAARSSSAHRGLAVSSP
CHWTFISPGHSGPHRTGFALAGO"
complement (<13180. .>13660)
/locus tag="OSJNBb0018B10.11"
/note="predicted by genemark hmm"
mRNA complement (join(<13180. .13414,13527. .>13660))
CDS /locus tag="OSJNBb0018B10.11"
/locus tag="OSJNBb0018B10.11"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAP54658.1"
/db_xref="GI:31433105"
/translation="MAIIMSVRNRRRAIGSEGGHLLGSBDTGKGLKILWGGAPILS
AGARICIVFSALJTSLAFLVPWRNPAIRPPVPAIAARVVRRESTVAVALVAVVN
ADESPRRHRRCCLLPRRR"
<14243. .>22346
/locus tag="OSJNBb0018B10.12"
/note="predicted by genescan"
mRNA join(<14243. .15321,16339. .16418,18176. .18376,19388. .19494,
20446. .20632,20837. .20875,22300. .>22346)
CDS /locus tag="OSJNBb0018B10.12"
/locus tag="OSJNBb0018B10.12"
/locus tag="OSJNBb0018B10.12"
/locus tag="OSJNBb0018B10.12"
/locus tag="OSJNBb0018B10.12"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAP54659.1"
/db_xref="GI:31433106"
/translation="MAGGGWSSLPADLLREVSGRILSYDADHLH:HQVCTHWRATLPV
SACRPVAVAGALQRLVITIGDYSRLRPGGAERMDAGLRAPPLRHCCHGASRWLA
LTDARSPTRALWEPLSGABEIPCLSPITRSLSGDPLASPGWVWVSQLKREGQ
KTLVCRGDAWMLFEHTGEIGSVFPHGKVYVIDIMSNIVCDLADGASAKCT
QFFVTRVSMLCSDKFHPBERGVHLVACDGEVLLVLRMGHPSLAEIYKPEKGNH
QLGGERVMDLGDHSLFVRGDTFSLSAKEPPAIKRNVCVYADKNYKQOHWISVPHLE
FPVAAILPWLAVAVRRPPRRPSPVSRHPPLFPFSLSLAAPGRATAPVVCRRJCAAT
PVAGKGRDGAGRALRPLSLPDSPRHERGLPAALRGVANSIEFPVLGRMAQDILTI
PASTVSLSEVSTSGRLVSFRSOUTPETTKALVCLQDWIRESGSSSELSTESIFDSK
VVLQVSFEAQSP"
<25677. .>28436
/locus tag="OSJNBb0018B10.13"
/note="similar tomato Cf 9 plant disease resistance
polyprotein GI:7489045 (Lycopersicon esculentum)"
mRNA join(<25677. .27248,27324. .>28436)
CDS /locus tag="OSJNBb0018B10.13"
/locus tag="OSJNBb0018B10.13"
/locus tag="OSJNBb0018B10.13"
/codon_start=1
/product="putative plant disease resistance polyprotein"
/protein_id="AAP54660.1"
/db_xref="GI:31433107"
/translation="MESKLEKLEGLIKMHIEDWLQSLSEKVTTKVDLAMDVFV
GVQQRCALRSMKRVANCSQAGILKSPGISQVSAGSSTPPLPPPPVPLVDGV
DLGRFGDGIYRRHQLPMDPFRFDSDVRIWLNMCETFYDMYQITQNFKYSAYVLH
MSGNAQWVSHYKLVNNEVNSDQFRMAVATEFVGVVEREKMSALDTLTQTFVTVYKQ
QFDLYVQIIRVEDPSPVGGKMLVTRFMNGLTEIRAAVVQLPDTVQSSAIAMQESV
LASRAIKALKKGFEVKLPQYKSDVSDDTTPGQVDKGLDWAKQLKDYRRAQGLCFKCGD
KYTEHRCVAVGGQIKAWQLEEVLTDDILDVIAAEQSDDEEDCHISINALTGAHLPTA
ILLRALVSHKVLILLVDYSSSFIDYNEAQMGLPLKPTPSLTVAVNGDCLPCQVY
VFGSWVIOGHFTYDMRVVITLGVSHDAIILQMDLQEWGMSCHWANKTLKFQYQGSWI
SLQGVQAAVPESEIAADVVHVIPSSQQLLTDVDFVHDKPCLPHRVEDHAILFLVP
DATPNCVPYRSPLOKDEIEKQVKEMIDAGLITPSMSPYASPVLLVKKDGSWRFCV
DYRLNALTIKSGFPMPPVDELLDELSTGTFKFKLDLKAGYHQIRMVETDEAKTAFKT
HGFQFQFVMPFGLTNAPSTFOCLMNSVFAPFIRKFLVFMNDILVYSPDLSDSHLHL
QQVFILRQHQJPAKSKTCTFACTQLELYGLYIISDKGVSTDPEKTAAMLAWPVTSVT
EURGFLGLTYRKRFGVGSYIARPUILLILKKKAFIWTDCCHAQKFNRISSRNCIKIP
VDPQGYTNDNDVDIQIHLVLTLLKYINAAEEKKONRS"
<30126. .>33383
/locus tag="OSJNBb0018B10.15"
/note="similar tomato Cf 9 plant disease resistance
polyprotein GI:7489045 (Lycopersicon esculentum)"
mRNA join(<30126. .30676,30790. .31002,31382. .>33383)
CDS /locus tag="OSJNBb0018B10.15"
/locus tag="OSJNBb0018B10.15"
/locus tag="OSJNBb0018B10.15"
/codon_start=1
/product="putative plant disease resistance polyprotein"
/protein_id="AAP54661.1"
/db_xref="GI:31433108"
/translation="MATARARVARTGDIRRRRRRKGVAKAADGGGDRRRRVARVA
TTHEGSKERKGEVLTGSDGRRRRGATEGGHAADGDRDDLKNEKKIKREGDRHR
RRPPBAGGGHGGPHRTMGALRRRTSTKEGTWMTSATRRRRRRRGGAAASLAAS
GFSGGGKGRHLHGAGMGSGDGGGEGGIRKSNPHLGRGRGRREGSGRGERGA
RAAWATPFGGGDVGTAAVAGAARRRRRREVGSGPDELGRRTSGCDDTATQLAFDQLK
Query Match 4.5%; Score 174.4; DB 8; Length 30029;
Best Local Similarity 49.9%; Pred. No. 1e-16;
Matches 696; Conservative 0; Mismatches 616; Indels 82; Gaps 7;
QY 1529 ACGTGCTCTCGGTGACGGCATCTTCAACGCCGACGGCGGAGCTGTGAGGAGAGCAGAGGA 1588
Db 295841 AGGACCTGTTTCGGGAGCGGATCTTCCCATCGACGGCGAGAAAGTGAAGCAGCAGAGGA 295782
QY 1589 AGACGGCGAGTTTCAGTTCGCTCCCAAGAACCTGAGGGATTTTCAGCGCCATTTGTGTCA 1648
Db 295781 AGATCGGCAGTAGACTTCTCCACAGGCGCCCTCCCGGACTTCAGCTGCGCGCTTCA 295722
QY 1649 GAGAGTACTCCTGAAGCTGTTCGGGTATATGAGCCAGGCATCCAAAGCGAGGCAAGTTG 1708
Db 295721 AGAGGAACGGCGCAAGCTGCGCGCATCGTCTCAACCAACCGCGCTCAACCAATCCA 295662
QY 1709 TGGACATGACGTGAGTCACTGCTC---CCTTCGCCATTGCCACATGAGCATTTCAACC 1765
Db 295661 TGGACTTCCAGGTAAATTAAGCGATCGTAATACAAATGCTAATTAATGCACGCGCGCG 295602
QY 1766 TGAGACACGAGAGTACTCCT-----TGCCGATTTCAGGAACCTTTACATGAGG 1810
Db 295601 GTGAGACGGTGAGTGACGTACGGTTTGTGTGCAAGCTTTGTTTCAAGGTTTGTATGCTGAG 295542
QY 1811 ATGACGTTGGACTCACTGCAAGGTTGGGTTGGGTTGGGTTGGGTTGGGTTGGGTTGGGTTG 1870
Db 295541 GCAACGATGGAAGTCCATCTTCCCATCGCATTCGGCAGACACCTCAACACCTCGACGCGC 295482
QY 1871 GATCTCCCGGAGACAGGTTTCGGCAGCGGTTTCGATCGCGCAACATCATCGTCAAGCTG 1930
Db 295481 TCCGGCGAGGGAGCGGCTTCGCGCGCGGTTTCGACGACGCGCAGTGTACCATGCTC 295422
QY 1931 CGGTTTCATGACCGCGCTGTGGCGGATCAAGAGGTTTCTTCCACGTCGGGTTCAGAGCCCTC 1990
Db 295421 CGTTACATCAGCCCGCTCTGAAGCTGGCAAGGCTCTCTCAACGTCGGGTTCAGAGCCATG 295362
QY 1991 CTAGCGCAGAGCATCAAGCTCGTGGACGAGTTTACCTACCGCTGATCCCGCGGGAGGAG 2050
Db 295361 CTCAGGAGAGGATCAAGTGTGTGACGCAATTCGTGTACAGGCTCATCCGTGCCAGGTCC 295302
QY 2051 GCCGAGATCGTCGAGGCCCGCGCCAGCGGCAAAAGGAGAGGTACGTGCATGACTGT 2110
Db 295301 GACGAGCTCTCAACTCACAGACTCCGTAAGCAACCCCTGAACTAAAGCAGCTCCTAC 295242
QY 2111 TTCGATTCTTCAGTTTCATCGTCTTTCGGCGGGATGGACCTGATCTCTGATTGATTATATC 2170

```

Db 295241 ACTCCGGTTTGACT-----GACATATATATAGATGATCATTTCTGGATGTGTAACTAAG 295187  
 QY 2171 CGTGTGACTTGTGAGGACAAATTAATGGGAGATGAAGACACGACATCCTCTGACGGTT 2230  
 Db 295186 C-----TGAATTTGTATCTGTACAGGTTTCGAGGAGGATATCTATCAAGATT 295137  
 QY 2231 CATCGAGCTAGGCGAGCGCGGAGACGACCGCGCGGCTTCGGGGAGCAGCAAGAGCTCGG 2290  
 Db 295136 TCTCCAGGCAACGA-----CCAGCATTTCTGTGTGCGATTACAGTACCTCGG 295089  
 QY 2291 GACCGTGTGCTCACTTCTGTATCGCGGGGGGACGACGCGGACGACGCGCTGCTG 2350  
 Db 295088 AGACATCACTACTGACATTTGTATAGCGGAGGACACACAGCTGGGGCGCTTGCTTG 295029  
 QY 2351 GTTACGACATGGCCATGTCACACCGGACGCTGGCGGAGAGCTGCGCGGAGCTGTG 2410  
 Db 295028 GTTCTGTATAGTGTGCAAGCACCGGAGTCCAGGAGAGATCTGCCAGAGCCAT 294969  
 QY 2411 CGGTTTCGAGGGAGCGCGCGGAGGAGGCGCTGCGCTGCTGCTGCTGCTGCGGCGCG 2470  
 Db 294968 GTTGGCCACAGCGCGCGGAGCACCGCTTCCGTCGACGAGTCTTTCGACAGCTTGA 294909  
 QY 2471 TGACGCGGACGACGAGCGCTGCGCGCGCGTGGCGGAGTTCGCGGCTCTCACTTA 2530  
 Db 294908 CCAAGCGCTGAACACATGACATCTGACGCGGACCTGACGAGACGCTCGGGTTTA 294849  
 QY 2531 CGACAGCTCGGCAAGCTGTCTACCTCCACGCGCTGCTGCTGCTGCTGCTGCTGCTG 2590  
 Db 294848 CCCTTCAGTTCCTCAATGTTGATTAATATGATCAGTTTCACTTA----- 294805  
 QY 2591 CCCCGCGCTCCCTCAGTGAGCGCGCGGACGACGCGACCTCCGTTCCAGACGACGATG 2650  
 Db 294804 -----CCATCGAGCTAAATTAATCAACGACATTTCTCTCATGAGTGGGTAAACAA 294754  
 QY 2651 CAGTGAGTGACCTGAATGCAATGCATGCACTTCCGCGCGGAGGACCCCAAGGGA 2710  
 Db 294753 CGTCTCTGTGATG-----ATGATCTTTTCTTTTCAGGAAACAGACAGT 294706  
 QY 2711 TCTTGAGGACGACGCTGCTGCTGCGGAGCGGACGAGGTGAGGCGCGGCGGATGGTACGT 2770  
 Db 294705 GCTTTTCAGACGATGATTGCCCCAAGGTTTAACTGACGAGGAGGACATCTGTTCT 294646  
 QY 2771 AGTGCCTTACTGATGGGCGGATGAGTAACTGAGGCGCGGCGGCGGAGCTTCC 2830  
 Db 294645 TCATCCCTACGATGAGGCGGATGAGAGCTTGTGGGCGAAAGACGCTGAATCTCC 294586  
 QY 2831 GGCGGAGCGGTGATCAACAGGATGGCGCTTCCGCAACGCGTCCGCTTCAAGTTCA 2890  
 Db 294585 GGCCTGAAAGTTGGCTCGAGAAACGCGCTCTTTCAGCGGAAAGCGGCTTCAAGTTTA 294526  
 QY 2891 CGCGCTTCCAGGCG 2904  
 Db 294525 CAGCATTTCAAGTG 294512

## RESULT 15

CNS08C9V

LOCUS

DEFINITION

Oryza sativa chromosome 12, . BAC OSJNBa004.1M09 of library OSJNBa from chromosome 12 of cultivar Nipponbare of ssp. japonica of Oryza sativa (rice), complete sequence.

ACCESSION

AL772417

VERSION

AL772417.3

KEYWORDS

HTG.

SOURCE

Oryza sativa (japonica cultivar-group)

ORGANISM

Oryza sativa (japonica cultivar-group)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartoideae; Oryzoideae; Oryza.

1 (bases 1 to 146519)

REFERENCE

Choisne, N., Orjeda, G., Cattolico, L., Demange, N., Wincker, P., Segurens, B., Pelletier, E., Scarpelli, C., Salanoubat, M.,

Weissenbach, J. and Quetier, F.  
 Oryza sativa chromosome 12 sequencing  
 Unpublished  
 2 (bases 1 to 146519)  
 Genoscope.  
 Direct Submission  
 Submitted (20-NOV-2003) Genoscope - Centre National de Sequencage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
 - Web : www.genoscope.cns.fr)  
 On Oct 2, 2002 this sequence version replaced gi:22003295.  
 Center: Genoscope / Centre National de Sequencage  
 Center code: GS  
 Web site: http://www.genoscope.cns.fr/  
 Contact: Seqref@genoscope.cns.fr

-----  
 The following sequence is oriented from the T7 to the SP6 end.  
 Upstream BAC (overlapping the T7 end) : OSJNBa0037B01 (AC=AL928751)  
 Downstream BAC (overlapping the SP6 end) : OSJNBa0068D20  
 (AC=AL731873) -----  
 FINISHED SEGMENT STARTS AT BASE 1  
 FINISHED SEGMENT ENDS AT BASE 146519  
 -----

## FEATURES

source

1. 146519  
 Location/Qualifiers  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="genomic DNA"  
 /cultivar="Nipponbare"  
 /sub\_species="japonica"  
 /db\_xref="taxon:39947"  
 /chromosome="12"  
 /clone="OSJNBa0041M09"  
 /clone\_lib="OSJNBa"

## ORIGIN

Query Match 4.2%; Score 162.4; DB 8; Length 146519;  
 Best Local Similarity 56.5%; Pred. No. 6.9e-15;  
 Matches 383; Conservative 0; Mismatches 286; Indels 9; Gaps 4;  
 QY 2399 CGCGAGCTGTGCGGTTGAGGCGGAGCGCGCGGCGGAGGCGCTCGCGCTGTGCC 2458  
 Db 8689 CTTGTTGTGCGGATGGCGCGCGGGGTACCCCGACGAGGCGATCTGCGACATGTGTGT 8748  
 QY 2459 CTGCGGCGGCTGAGCGCGACGACGAGGCTTCGCCCGCGCGGCGAGCTTCGCGG 2518  
 Db 8749 GAGCTTCATCATGCGCGCGCGGACACGACGCTGCTGCGGCGC---TGACGTGTTCTTCTG 8805  
 QY 2519 CTTCTCTACCTACGACAGCTCGGCAAGCTGCTTCTACCTCCACGCTGCTCACCGAGAC 2578  
 Db 8806 GCTCATGACGCCACCGCGAGCTGGAGCGGAGGTTCTTGGACGAGGTGACGCGTGCAT 8865  
 QY 2579 GCTCGGCTGTATCCCGCGCTCCCTCAGTGTAGGCGGCGCGGACACGCGACCTCCGCTCA 2638  
 Db 8866 GGGCGACGCGCGGATGGCGGTGGCTCTCGCGCGCTGGAC--CTCAGGCGCTCGCGCAG 8923  
 QY 2639 GAGCACAGCATGCACTGAGTGACCTGATGCAATGCAATGCAATGCAATGCAATGCAATG 2697  
 Db 8924 GCGGCGTCTCCAGCGCGCTCTGCGAGAGATGCGGCTGTACCCGCGGTGGGCTGG 8983  
 QY 2698 GACCCCAAGGGATCCTGAGGAGCAGCTGCTGCCGAGCGGACGAGGAGTGAAGGCGCGC 2757  
 Db 8984 GACTCGAAGCAGCGCGGAGGACGACGCTGCTCCCGGACGCGACGCGCGTGGGCGCGCG 9043  
 QY 2758 GGGATGTTGACGTAGTGCCTTCTGATGGGCGGATGAGTACAACTTGGGCGCGCGAC 2817  
 Db 9044 GACCGCGTCACTTCTCCCTTACGGGATGGGAGATGGAGGCGCATCTTGGGCGCGCGAC 9103  
 QY 2818 GCGGCGAGCTTCCGCGCGGAGCGGTGGATCAACAGGATGGCGCGTTCGCGCAACGCGT 2877  
 Db 9104 GCGGCGAGTTACGCGCGGAGGTGGCTTGGCGCGCGCGCGCGCGCGCGCGCGCTGTG 9163  
 QY 2878 CGTTCAAGTTCAAGCGGCTTCAGGCGGCGCGGAGATCTGCTGGGCAAGGACTCGGCG 2937  
 Db 9164 CGTTCAAGTACCGGTTTCCAGGCGGCGCGGACGCTGCTCGGCAAGGACATGGCG 9223

QY	2938	TACCTGCAGATGAG---	ATGGCGCTGGCCATCCTTTGCGCTTCTACAGCTTCCGGCTG	2994
Db	9224	TTCTGTCAGATGAAGTT	CGTGGCCAGCGCCGTCTCCGGCGGTTGAGCTCCGCCCGGTG	9283
QY	2995	CTGAGGGGGACCCGGTGCAGTAC	CGCATGATGACCATCCTCTCCATGCGCGCACGGCCTC	3054
Db	9284	GCGCCGAGGGGCTCGCCGAGTT	CGTGGCCGTGCTCACCGCGCACATGGCCGGGGGCTC	9343
QY	3055	AAGTCCCGCTCTTAGG	3072	
Db	9344	AAGGTGACGGTGAGGAGG	9361	

Search completed: March 7, 2004, 15:19:58  
 Job time : 9799 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 5, 2004, 22:02:35 ; Search time 5017 Seconds  
(without alignments)

11344.891 Million cell updates/sec

Title: US-10-021-657-1

Perfect score: 1906

Sequence: 1 gaattcgccagcagggaagc.....aaaaaaaaaaaaaactcgag 1906

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_htc:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_hcc:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: em\_gss\_hum:\*

18: em\_gss\_inv:\*

19: em\_gss\_pin:\*

20: em\_gss\_vrt:\*

21: em\_gss\_fur:\*

22: em\_gss\_nam:\*

23: em\_gss\_mus:\*

24: em\_gss\_pro:\*

25: em\_gss\_rod:\*

26: em\_gss\_phg:\*

27: em\_gss\_vrl:\*

28: gb\_gsl:\*

29: gb\_gsl2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	818.4	42.9	997	11	AY110120 Zea mays
C 2	651.2	34.2	1002	29	CG344416 OGLDH18TV
C 3	619.6	32.5	887	29	CC611162 OGLDH18TV
C 4	579	30.4	833	29	CG317324 OG0CS18TH

5	573.4	30.1	720	13	BQ840694
6	569	29.9	859	29	CG317335
7	554	29.1	563	10	AW519943
8	540	28.3	880	29	CG047034
9	533.8	28.0	578	12	BI075441
10	527.4	27.7	879	29	CC685050
11	509.4	26.7	585	12	BI075273
12	482	25.3	606	10	BE494080
13	473	24.8	748	28	BZ532472
14	463.6	24.3	591	10	AW424821
15	461.4	24.2	517	13	BQ841568
16	457.4	24.0	681	29	CG317414
17	454.2	23.8	819	28	CC368220
18	454.2	23.8	882	29	CC635363
19	454.2	23.8	991	29	CG047035
20	448.2	23.5	518	12	BG948686
21	445.6	23.4	620	13	BU996440
22	441.6	23.2	625	13	BQ840748
23	438.2	23.0	864	29	CG344408
24	427.4	22.4	773	28	CC368222
25	427.4	22.4	887	29	CG067339
26	420.4	22.1	458	12	BI246000
27	417.4	21.9	509	12	BI099541
28	409.6	21.5	470	12	BI246162
29	406.6	21.3	615	29	CG611171
30	394.4	20.7	649	12	BG948366
31	391.4	20.5	514	10	BF292135
32	388.2	20.4	408	12	BM500648
33	368.8	19.3	469	12	BG274193
34	361	18.9	535	10	BF292000
35	344.2	18.1	406	12	BI141660
36	338.2	17.7	400	12	BI141132
37	330.2	17.3	848	29	CC723851
38	330	17.3	712	29	CG067338
39	324.4	17.0	679	14	CD491751
40	308.2	16.2	383	12	BG274260
41	285	15.0	645	9	AU239680
42	284.2	14.9	794	29	CC685043
43	274	14.4	534	29	CG273559
44	272	14.3	526	12	BI074975
45	271	14.2	477	12	BI245608

## ALIGNMENTS

RESULT 1	AY110120/c	997 bp	mRNA	linear	HTC 17-OCT-2002
LOCUS	AY110120				
DEFINITION	Zea mays CL58113.1 mRNA sequence.				
ACCESSION	AY110120				
VERSION	AY110120.1	GI:21214224			
KEYWORDS	HTC.				
SOURCE	Zea mays				
ORGANISM	Zea mays				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.				
AUTHORS	Haney,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S., Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.				
TITLE	Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes				
JOURNAL	Unpublished (2002)				
REFERENCE	2 (bases 1 to 997)				
AUTHORS	Coe,E.H.				
JOURNAL	Direct Submission				
TITLE	Submitted (25-Apr-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA				
COMMENT	If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR, the www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the				



Db 936 TCACAAAGTACATCGCGCTCCTTCTGCTGTCCTCTCATGGATCCTGCTCCAGAGGTGGAG 877  
 QY 129 CCTGAGGAAGCAAGAGCCCGAGATCATGGCCAGTTCATCGGCGCAACGGTGGAGCAGCT 188  
 Db 876 CCTGAGGAAGCAAGAGCCCGAGATCATGGCCAGTTCATCGGCGCAACGGTGGAGCAGCT 817  
 QY 189 GAGGAATCTACACCGGATGACGACCTGGTGTGCGGTACCTGTACCGGCACAGGACAGT 248  
 Db 816 GAGGAATCTACACCGGATGACGACCTGGTGTGCGGTACCTGTACCGGCACAGGACAGT 757  
 QY 249 GACGTCGACATGCCGTTCACTTCTACACTACCTACCTGCTGACCGGTGAATGTCGAGCA 308  
 Db 756 GACGTCGACATGCCGTTCACTTCTACACTACCTACCTGCTGACCGGTGAATGTCGAGCA 697  
 QY 309 TGTCTCTCAAGACTAACTTCCACCAATACCCCA-----AGGAA 346  
 Db 636 CAGTCCTTCGGAATCAGAGCTGAAGCTGAATCGAATGTGCTGAACACCGTGTAGGAA 577  
 QY 347 TCGTGTACAGATCCTATCATGAGCTGCTCCTCGGTGACGGCATCTTCAACGCGACGCG 406  
 Db 576 TCGTGTACAGATCCTATCATGAGCTGCTCCTCGGTGACGGCATCTTCAACGCGACGCG 517  
 QY 407 AGCTGTGAGGAAGCAGAGGAGACGCGAGTTCGAGTTTCGCTCTCAAGAACCTGAGGG 466  
 Db 516 AGCTGTGAGGAAGCAGAGGAGACGCGAGTTCGAGTTTCGCTCTCAAGAACCTGAGGG 457  
 QY 467 ATTTTCAGCGCCATGTGTTCAGAGATCTCCCTGAAGCTGTGCGGTATCTAGACGAGG 526  
 Db 456 ATTTTCAGCGCCATGTGTTCAGAGATCTCCCTGAAGCTGTGCGGTATCTAGACGAGG 397  
 QY 527 CATCAAGGCGAGCAAGATGTCGACATG-----555  
 Db 396 CATCAAGGCGAGCAAGATGTCGACATGCAAGGTGAGATATCATGTTCCCTTCCCATG 337  
 QY 556 -----CAGGAACCTTTACA 568  
 Db 336 CCAACATGAGCATTTCAACCTGAGACAGAGCTACCTTCCCGATTTCAGGAACCTTTACA 277  
 QY 569 TGAGATGACGTGACATCTCATCTGCAAGTGTGGTTCGGGTGAGATCGGACGCTGT 628  
 Db 276 TGAGATGACGTGACATCTCATCTGCAAGTGTGGTTCGGGTGAGATCGGACGCTGT 217  
 QY 629 CGCAGATCTCCCGAGAACAGCTTCGCGCAGGCGTTCGATGCGCGCAACATCATCA 688  
 Db 216 CGCAGATCTCCCGAGAACAGCTTCGCGCAGGCGTTCGATGCGCGCAACATCATCA 157  
 QY 689 CGCTCGGTTTCATCGACCGCTGTGGCGCATCAAGAGGTTCTTCCACGTTCGGGTGAGG 748  
 Db 156 CGCTCGGTTTCATCGACCGCTGTGGCGCATCAAGAGGTTCTTCCACGTTCGGGTGAGG 97  
 QY 749 CCCTCCTAGCGAGAGCATCAAGCTCTGTGGAGAGTTCACCTACAGCGTGTCCGCGCA 808  
 Db 96 CCCTCCTAGCGAGAGCATCAAGCTCTGTGGAGAGTTCACCTACAGCGTGTCCGCGCA 37  
 QY 809 GGAAGCGCGAGATCTGAGGTCGCGGCGCAGCGCA 844  
 Db 36 GGAAGCGCGAGATCTGAGGTCGCGGCGCAGCGCA 1

RESULT 3  
 LOCUS CC611162/c  
 DEFINITION OGLEB88TH ZM\_0.7\_1.5\_KB Zea mays genomic clone ZMMBMA0315P08,  
 genomic survey sequence.  
 ACCESSION CC611162  
 VERSION CC611162.1 GI:31972583  
 KEYWORDS GSS.  
 SOURCE Zea mays  
 ORGANISM Zea mays

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD  
 clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 887)

Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,  
 Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,  
 Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.

Consortium for Maize Genomics

Unpublished (2002)

Other GSSs: OGLEB88TV

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TR

Class: sheared ends.

Location/Qualifiers

source

1..887  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /strain="B73"  
 /db\_xref="taxon:4577"  
 /clone="ZMMBMA0315P08"  
 /clone\_lib="ZM\_0.7\_1.5\_KB"  
 /notes=vector: pBSK-; Site 1: HincII; 0.7-1.5 kb  
 methylation filtered genomic DNA library"

## ORIGIN

Query Match 32.5%; Score 619.6; DB 29; Length 887;  
 Best Local Similarity 97.1%; Pred. No. 9e-71;  
 Matches 631; Conservative 0; Mismatches 19; Indels 0; Gaps 0;  
 QY 1232 TCGCGCTGTACCCCGCGCTCCCTCAGACCCCGAGGGATCTCTGGAGACGACGTGTGTC 1291  
 Db 851 TGCACATGCACTTTCGCGCGCGCAGGACCCCAAGGGGATCTCTGGAGACGACGTGTGTC 792  
 QY 1292 CGGACGGGACGAGGTGAGGCGCGGGGATGTGACGTAGTGCCTTCTCGATGGGC 1351  
 Db 791 CGGACGGGACGAGGTGAGGCGCGGGGATGTGACGTAGTGCCTTCTCGATGGGC 732  
 QY 1352 GGATGGAGTACAATGGGCGCCCGACCGCGGAGCTTCCGCGCGAGCGGTTGATCAACG 1411  
 Db 731 GGATGGAGTACAATGGGCGCCCGACCGCGGAGCTTCCGCGCGAGCGGTTGATCAACG 672  
 QY 1412 AGGATGGCGGTTCCGCAACGCGTTCGCGGTTCAGAGTTCA CGGCGTTCCAGCGGGCGCA 1471  
 Db 671 AGGATGGCGGTTCCGCAACGCGTTCGCGGTTCAGAGTTCA CGGCGTTCCAGCGGGCGCA 612  
 QY 1472 GGATTCGCTGGGCAAGGACTCGGCGGTACCTGCAGATGAAGATGGCGTGGCCATCTCT 1531  
 Db 611 GGATTCGCTGGGCAAGGACTCGGCGGTACCTGCAGATGAAGATGGCGTGGCCATCTCT 552  
 QY 1532 TCCGCTTCTACAGCTTCGCGGTCTGGAGGGGCAACCGGTGCAGTACCGCATGATGACCA 1591  
 Db 551 TCCGCTTCTACAGCTTCGCGGTCTGGAGGGGCAACCGGTGCAGTACCGCATGATGACCA 492  
 QY 1592 TCCTCTCCATGGCGACGGCTCAAGGTTCGCGTCTCTAGGGCGGTCTGATGTCATGCGG 1651  
 Db 491 TCCTCTCCATGGCGACGGCTCAAGGTTCGCGTCTCTAGGGCGGTCTGATGTCATGCGG 432  
 QY 1652 ATTTGGATATGGATATGTCGCGCTTAATCCAGACAAATAACCGTCTGTTTACAAATTT 1711  
 Db 431 ATTTGGATATGGATATGTCGCGCTTAATCCAGACAAATAACCGTCTGTTTACAAATTT 372  
 QY 1712 GCATGCATGATGTAAGGAAAGCATGGCTTTCATTCGTCGCTTGGCTTAAGCCCTAAA 1771  
 Db 371 GCATGCATGATGTAAGGAAAGCATGGCTTTCATTCGTCGCTTGGCTTAAGCCCTAAA 312  
 QY 1772 AACTCCGTGGGTCTTTCGGAACCAACCATCATCACTAGTCTTTGTACTTACTCTCAGTG 1831  
 Db 311 AACTCCGTGGGTCTTTCGGAACCAACCATCATCACTAGTCTTTGTACTTACTCTCAGTG 252





```

484 CCTCGGGACGTGTGTCTCAACTTCGTGATCGCGCGGACACGACGCGGACGCGCT 543
996 GTCTGTGTTACGACATGCGCATGTCCACCGGAGCTGTGCGCGGAGCTGCGCGCGA 1055
544 GTCTGTGTTACGACATGCGCATGTCCACCGGAGCTGTGCGCGGAGCTGCGCGCGA 603
1056 GCTGTGCGCGTTCGAGCGGAGCGCGCGCGGAGGCGTCAAGCTGCTGCTGCTGCGG 1115
604 GCTGTGCGCGTTCGAGCGGAGCGCGCGCGGAGGCGTCAAGCTGCTGCTGCGG 663
1116 CGGCGCTGACCGCGGACGACAAAGCGGTTTCGCGCGCGGCGCGGAGGCGGCGCTCT 1175
664 CGGCGCTGACCGCGGACGACAAAGCGGTTTCGCGCGCGGCGCGGAGGCGGCGCTCT 723
1176 CACCTACGACGCTTCGCGGAGCTGCTTACTCTCAGCGCTGCGTCAACGAGAGCGCTCG 1235
724 CACCTACGACGCTTCGCGGAGCTGCTTACTCTCAGCGCTGCGTCAACGAGAGCTCG 783
1236 CTTGTACCGCGCGCTGCTCAGG 1258
784 CTTGTACCGCGCGCTCCTCAGG 806

RESULT 7
AW519943
LOCUS 660053C10.Y1 660 - Mixed stages of anther and pollen Zea mays cDNA,
DEFINITION mRNA sequence.
ACCESSION AW519943
VERSION AW519943.1 GI:7162216
KEYWORDS EST.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 563)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
JOURNAL University
COMMENT Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 660053 row: C column: 10.
Location/Qualifiers
1. 563
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="Ohio43"
/db_xref="taxon:4577"
/tissue_type="whole premeiotic anthers to pollen shed"
/dev_stage="premeiotic anthers to pollen shed"
/lab_host="XLOLR"
/clone_lib="660 - Mixed stages of anther and pollen"
/notes="Organ: anthers; Vector: Lambda Zap; Site_1: EcoRI;
Site_2: XhoI; Anther and pollen cDNA library.
Directionally sequenced with 5' end at the EcoRI site.
Created by Amie Franklin."

FEATURES
source
Query Match 29.1%; Score 554; DB 10; Length 563;
Best Local Similarity 99.1%; Pred. No. 3e-62;
Matches 557; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

914 GCCTCTCGGGACGATAGAGAGCTCTCGGACCTGTGTGCTCAACTTCTGTGATCGCGGGC 973
62 GCCTCTCGGGACGACAAAGAGCTCTCGGACCTGTGTGCTCAACTTCTGTGATCGCGGGC 121
974 GGCACACGCGGCGGACGAGCTCTGCTGTTTACGACATGCGCATGTCCACCCCGGAGC 1033
122 GGCACACGCGGCGGACGAGCTCTGCTGTTTACGACATGCGCATGTCCACCCCGGAGC 181
1034 TGGCCGAGAAAGCTTCCGCGGAGCTGTGCGGCTTTCGAGCGGAGCGCGCGGAGGAGG 1093
182 TGGCCGAGAAAGCTTGGCGCGGAGCTGTGCGGCTTTCGAGCGGAGCGCGCGGAGGAGG 241
1094 GCGTCACGCTGCTGCTCTGCGGCGGCGCTGACGCGGAGCAAGCGGTTTGGCGCGCGCG 1153
242 GCGTCGCGCTCTGCTCTGCGGCGGCGCTGACGCGGAGCAAGCGGTTTGGCGCGCGCG 301
1154 TGGCGCAGTTTCGCGGCGCTCTCCTACCTACGACAGCTCTCGGCAAGCTGCTTACCTCCACG 1213
302 TGGCGCAGTTTCGCGGCGCTCTCCTACCTACGACAGCTCTCGGCAAGCTGCTTACCTCCACG 361
1214 CTTGCTGTCACCGAGAGCTCTCGGCTGTACCCCGCGCTTCCCTCAGGACCCCAAGGGATCC 1273
362 CTTGCTGTCACCGAGAGCTCTCGGCTGTACCCCGCGCTTCCCTCAGGACCCCAAGGGATCC 421
1274 TGGAGGACGAGCTGCTCTCGGAGCGGAGAGCTGAGGCGGCGGCGGATGGTGACGTACG 1333
422 TGGAGGACGAGCTGCTCTCGGAGCGGAGAGCTGAGGCGGCGGCGGATGGTGACGTACG 481
1334 TGCCTACTCTGATGGCGGATGAGTACAACTGGGCGCGCGCGCGGAGCTTCCGCGC 1393
482 TGCCTACTCTGATGGCGGATGAGTACAACTGGGCGCGCGCGCGGAGCTTCCGCGC 541
1394 CGGAGCGGTGGATCAACGAGGA 1415
542 CGGAGCGGTGGATCAACGAGGA 563

RESULT 8
CG047034/c
LOCUS CG047034 880 bp DNA linear GSS 19-AUG-2003
DEFINITION PUFWN66TB ZM 0.6 1.0 KB Zea mays genomic clone ZMMBta0749K12,
genomic survey sequence.
ACCESSION CG047034
VERSION CG047034.1 GI:33919214
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 880)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
Bennetzen,J.
TITLE Maize Genomics Consortium
JOURNAL Unpublished (2003)
COMMENT Other GSSs: PUFWN66TD
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0206
Email: whitelaw@tigr.org
Seq primer: TR
Class: sheared ends.
Location/Qualifiers
1. 880
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBta0749K12"

914 AGATGAGCAGCAGCATCTGTCACGTTTCATCGAGCTGGCGGAGGCGCGGAGCGCGG 913
2 AGATGAGCAGCAGCATCTGTCACGTTTCATCGAGCTGGCGGAGGCGCGGAGCGCGG 61

```



```

365 AGGAGAAGATGAAGACGACATCTGTACGGTTTCATCGAGCTGGCGGAGCGCGGACG 424
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

```

```

699 CCAATTACCCCAAGGTAATGACCTGAACCTCACTGATGTTCACTTTCGGAATCAGAGC 640
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

```

```

RESULT 11
BI075273
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

```

BI075273 585 bp mRNA linear EST 20-JUN-2001
IP1_18_C12.b1_A002 Immature pannicle 1 (IPI) Sorghum bicolor cDNA,
mRNA sequence.
BI075273.1 GI:14513930
EST.
Sorghum bicolor (sorghum)
Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 585)
Klein, R.R., Cordonnier-Pratt, M.-M., Gingle, A., Sudman, M. and
Pratt, J.H.
An EST database from Sorghum: developing preanthesis pannicles
Unpublished (2001)
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions

```

below Phred quality 16. The threshold for high quality sequence is 20. Three-prime sequences, which are obtained with PolyTmix or T7 sequencing primer, are presented as the reverse complement.

Seq primer: JEN REV

High quality sequence stop: 519

POLYA=No.

Location/Qualifiers

1. 585

/organism="Sorghum bicolor"

/mol\_type="mRNA"

/cultivar="BTx623"

/db\_xref="taxon:4558"

/clone\_lib="Immature pannicle 1 (IP1)"

/note="Organ: Developing preanthesis panicles; Vector:

pBluescript II SK(-) from Lambda Zap II; Site 1: XhoI;

Site 2: EcoRI; The library was made from poly-A RNA in the

cloning vector lambda Zap II. Clones to be sequenced were

prepared by mass excision."

## FEATURES

source

Query Match 26.7%; Score 509.4; DB 12; Length 585;

Best Local Similarity 93.6%; Pred. No. 1.7e-56;

Matches 544; Conservative 0; Mismatches 31; Indels 6; Gaps 1;

Qy 15 GGAAGCTCACCTCAGCGCGGCGAGCCATCGCCATTCCTCCCATAGCAGGGCCTCAAA 74

Db 10 GGAAGCTCACCTCAGCGCGGCGAGCCATTCCTCCCATAGCAGGGCCTCAAA 63

Qy 75 GTACATCGCGCTCTCTTCGTGTTGCTCTCTCATGATCTGTGTCAGAGGTGAGCCTGAG 134

Db 64 GTACATCGCGCTCTCTTCGTGTTGCTCTCTCATGATCTGTGTCAGAGGTGAGCCTGAG 123

Qy 135 GAAGCAGAAAGCCGCGAGATCATGGCCAGTCATCGGCCAACGGTGGAGCAGCTGAGGAA 194

Db 124 GAAGCAGAAAGCCGCGAGATCATGGCCAGTCATGGGCCAACGGTGGAGCAGCTGAGGAA 183

Qy 195 CTACCACCGATGACGACTGGCTTGTCCGGTACTGTCACGACAGGACAGGACAGTACCGT 254

Db 184 CTACCACCGATGACGACTGGCTTGTCCGGTACTGTCACGACAGGACAGGACAGTACCGT 243

Qy 255 CGACATGCCGTCTACTTCTACACTACATGCTGACCCGGTGAATGTCGAGCATGTCTCT 314

Db 244 CGACATGCCGTCTACTTCTACACTACATGCTGACCCGGTGAATGTCGAGCATGTCTCT 303

Qy 315 CAAGACTAACTTCAACAATACCCAGGGAATCGGTACAGATCCTTACATGGACGCTGCT 374

Db 304 CAAGACTAACTTCAACAATACCCAGGGAATCGGTACAGATCCTTACATGGATGCTGCT 363

Qy 375 CTTGGTGAAGCATCTTCAAGCGCGAGCGGAGTGTGGAGGAAGCAGAGAACGCGC 434

Db 364 CTTGGTGAAGCATCTTCAAGCGCGAGCGGAGTGTGGAGGAAGCAGAGAACGCGC 423

Qy 435 GAGTTTCAGTTCGCTCCAGAACTGAGGATTCAGGCCATTTGTTTCAGAGATGA 494

Db 424 GAGTTTCAGTTCGCTCCAGAACTGAGGATTCAGGCCATTTGTTTCAGAGATGA 483

Qy 495 CTCCTCTGAAGCTGTCCGGTATTAAGCAGGAGCATCCAAAGCAGGACAAAGTTGTGACAT 554

Db 484 CTCCTCTGAAGCTGTCCGGTATTAAGCAGGAGCATCCAAAGCAGGACAAAGTTGTGACAT 543

Qy 555 GCAGGAACCTTTACATGAGGATGACGCTGACATCTTGCA 595

Db 544 GCAGGAACCTTTACATGAGGATGACATCTTGCA 584

## RESULT 12

BE494080

LOCUS

DEFINITION

WHE1277\_F02\_L03ZS Secale cereale anther cDNA library

SECALCERE

ACCESSION

BE494080

VERSION

BE494080.1

GI:9660673

KEYWORDS

606 bp mRNA linear EST 02-AUG-2000  
WHE1277\_F02\_L03ZS Secale cereale anther cDNA library Secale cereale  
cDNA clone WHE1277\_F02\_L03, mRNA sequence.

## SOURCE

ORGANISM

Secale cereale (rye)

Secale cereale

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Poideae; Triticeae; Secale.

1 (bases 1 to 606)

Anderson, O.D., Butler, E., Chao, S., Choi, D.W., Close, T.J.,

Fenton, R.D., Gustafson, J.P., Han, P.S., Hsia, C.C., Kang, Y.,

Lazo, G.R., Miller, R., Rausch, C.J., Ross, K., Seaton, C.L., and

Tong, J.C.

The structure and function of the expressed portion of the wheat

genomes - Anther cDNA library from rye

Unpublished (2000)

Contact: Olin Anderson

US Department of Agriculture, Agriculture Research Service, Pacific

West Area, Western Regional Research Center

800 Buchanan Street, Albany, CA 94710, USA

Tel: 5105595773

Fax: 5105595818

Email: andersowp.usda.gov

Sequence have been trimmed to remove vector sequence and low

quality sequence with phred score less than 20

Seq primer: Stratagene SK primer.

Location/Qualifiers

1. 606

/organism="Secale cereale"

/mol\_type="mRNA"

/cultivar="Blanco"

/db\_xref="taxon:4550"

/clone="WHE1277\_F02\_L03"

/tissue\_type="Anther"

/dev\_stage="Adult plant before anthesis"

/lab\_host="E. coli SOLR"

/clone\_lib="Secale cereale anther cDNA library"

/note="Vector: Lambda Uni-ZAP XR, excised phagemid;

Site 1: EcoRI; Site 2: XhoI; plants were grown in the

greenhouse. Anthers were harvested and pooled from early

meiosis to late meiosis. The tissue, total RNA, and

poly(A) RNA were prepared (Butler, Ross and Gustafson) at

University of Missouri, Columbia. A cDNA library was

made, and the cDNA clones were in vivo excised to give

pBluescript phagemids in the TJ Close lab (Choi, Close,

Fenton) at the University of California, Riverside.

Plasmid DNA preparations and DNA sequencing were performed

in the OD Anderson lab (all other authors)."

ORIGIN

Query Match 25.3%; Score 482; DB 10; Length 606;

Best Local Similarity 87.5%; Pred. No. 6e-53;

Matches 527; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

Qy 988 ACAGCGCTGTCGTGTTACGACATGGCCATGTCCACCCGAGCTGGCCGAGAGCTG 1047

Db 5 ACAGCGCTGTCGTGTTACGACATGGCCATGTCCACCCGAGCTGGCCGAGAGCTG 64

Qy 1048 CGCGCGAGCTGTGCGGCTTCGAGCGGAGCGCGCGGAGGAGGCGTCCAGCTCGT 1107

Db 65 CGCGCGAGCTGTGCGGCTTCGAGCGGAGCGCGCGGAGGAGGCGTCCAGCTCGT 124

Qy 1108 CTCTGCGCGCGCTGTGACCGGACGACAAAGCGGTTTCGCGCGCGCGCTGGCGAGTTCCGCG 1167

Db 125 CCCTGACGAGCGGAGGGGCTTCGACGAGGCGCTTCGCGCGCGCGCTGGCGAGTTCCGCG 184

Qy 1168 GGCCTCTCTACCTACGACAGCTCGGCAAGTGTGTACTCTCCAGCCTGGTGCACCGAG 1227

Db 185 GGCCTCTCTACCTACGAGCGGCTCGGCAAGTGTGTACTCTCCAGCCTGGTGCACCGAG 244

Qy 1228 AGCTTCCGCTGTATCCCGCGCGCTTCCTCAGACCCAGGGAGTCTCTGAGGAGCAGCTG 1287

Db 245 AGCTTCCGCTGTATCCCGCGCGCTTCCTCAGAGACCCAGGGAGTCTCTGAGGAGCAGCTG 304

Qy 1288 CTGCGGAGCGGACGAAAGTGTGAGGCGCGCGGAGTGTGAGTACGTGCTCCCTTCTCGATG 1347

Db 305 CTCCTCCGAGCGGACCAAGGTGCGCGCGGGGATGCTGACGTACTGCTCCATG 364

QY 1348 GGGCGGATGAGTACAACTGGGCCCCCGACGGCGGAGCTTCGGCCGAGCGGTGATC 1407

Db 365 GGGCGGATGAGTACAACTGGGCCCCCGACGGCGGAGCTTCGGCCGAGCGGTGATC 424

QY 1408 AACGAGATGGCGGCTTCGCAACGGGTGCGCGCTTCAAGTTCAAGCTTCCAGGGCGG 1467

Db 425 GGGGACGACGGCGGCTTCGCAACGGGTGCGCGCTTCAAGTTCAAGCTTCCAGGGCGG 484

QY 1468 CCGAGATCTGCTGGGCAAGGACTCGGCTACTGTCAGATGAAGTGGCGTGGCCATC 1527

Db 485 CCGGATCTGCTGGCAAGGACTCGGCTACTGTCAGATGAAGTGGCGTGGCCATC 544

QY 1528 CTCTCCGCTTCTACAGTTCGCGCTGCTGGAGGGGACCCGGTGCAGTACCCATGATG 1587

Db 545 CTGTGCAAGTCTTCTAGGTTGAGCTGCTGGAGGGCCACCCGCTCAAGTACCCATGATG 604

QY 1588 AC 1589

Db 605 AC 606

RESULT 13

BZ532472

LOCUS OGA1V02TC ZM2 0.7-1.5 KB Zea mays DNA linear GSS 16-DEC-2002

DEFINITION genomic survey sequence.

ACCESSION BZ532472

VERSION BZ532472.1 GI:27077158

KEYWORDS GSS.

SOURCE Zea mays

ORGANISM Zea mays

REFERENCE Rukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.

AUTHORS Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T., Resnick, A., Fraser, C.M., Sudman, M.A., Bedell, J.A., Rohlfing, T., Citek, R.W., Nummer, A., Robbins, D. and Lakey, N.

TITLE Consortium for Maize Genomics

JOURNAL Unpublished (2002)

COMMENT Contact: Cathy Whitelaw

TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TF

Class: sheared ends.

FEATURES

source

1..748

/organism="Zea mays"

/mol\_type="genomic DNA"

/strain="B73"

/db\_xref="taxon:4577"

/clone="ZMMEMA0071A04"

/clone\_lib="ZM2 0.7-1.5 KB"

/note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb methylation filtered genomic DNA library"

ORIGIN

Query Match 24.8%; Score 473; DB 28; Length 748;

Best Local Similarity 82.7%; Pred. No. 8.3e-52;

Matches 608; Conservative 0; Mismatches 5; Indels 122; Gaps 1;

QY 845 AACGAGAGATGAAGACGACATCTCTACGGTTCATCGAGCTGGCGAGGCGCGG 904

Db 14 AAAATGGGAGATGAAGACGACATCTCTACGGTTCATCGAGCTGGCGAGGCGCGG 73

QY 905 ACGAGCGCGCGGCTTCGGGGACGATAGAGCCTCCGGGACGCTGGTGTCACTTCGTA 964

Db 74 ACGAGCGCGCGGCTTCGGGGACGATAGAGCCTCCGGGACGCTGGTGTCACTTCGTA 133

QY 965 TCGCGGCGGACACGACGCGGAGCAAGCTGTCTGTGTTCACGCACATGGCCATGCCC 1024

Db 134 TCGCGGCGGACACGACGCGGAGCAAGCTGTCTGTGTTCACGCACATGGCCATGCCC 193

QY 1025 ACCCGGAGCTGGCCGAGAACTGCGCCCGAGCTGTGCGCGTTCAGGCGGAGCGCGCG 1084

Db 194 ACCCGGAGCTGGCCGAGAACTGCGCCCGAGCTGTGCGCGTTCAGGCGGAGCGCGCG 253

QY 1085 GCGAGGCGGCTGACGCTGCTGCTGCGCGCGCTGACGCCGACGACAAAGGCGTTCG 1144

Db 254 GCGAGGCGGCTGACGCTGCTGCTGCGCGCGCTGACGCCGACGACAAAGGCGTTCG 313

QY 1145 CCGCGCGCTGCGGAGTTTCGCGGCGCTCTCACTACGACAGCCTTCGCAAGTGGTCT 1204

Db 314 CCGCGCGCTGCGGAGTTTCGCGGCGCTCTCACTACGACAGCCTTCGCAAGTGGTCT 373

QY 1205 ACTTCACGCTGGGTACCGAGAGCTGCGCGCTGACCGCGCGCTCCCT----- 1254

Db 374 ACTTCACGCTGGGTACCGAGAGCTGCGCGCTGACCGCGCGCTCCCT----- 433

QY 1255 ----- 1254

Db 434 CGCCCGTCACTACGACCTCGGTCGCGATGCAACGCAATATGTGGTGTCCGAGAGC 493

QY 1255 -----CAGGACCC 1262

Db 494 ACAGCATGCACTGAGTGGACCTGAATGCAATGCACATGTCACCTTGGCGCGCGCAGGACCC 553

QY 1263 CAAGGGGATCTGGAGAGCAAGTGTCTGCGGACGGGAGGAGGTGAGGCGCGCGGAT 1322

Db 554 CAAGGGGATCTGGAGAGCAAGTGTCTGCGGACGGGAGGAGGTGAGGCGCGCGGAT 613

QY 1323 GGTGAGTACTGCTCCCTACTCTGATGGGCGGATGAGTACAACTGGGCGCGCGCAGCGGC 1382

Db 614 GGTGAGTACTGCTCCCTACTCTGATGGGCGGATGAGTACAACTGGGCGCGCGCAGCGGC 673

QY 1383 GAGCTTCCGCGCGGAGCGGTGGATCAACAGGATGCGGCTTCGCAACGCGTCGCGGT 1442

Db 674 GAGCTTCCGCGCGGAGCGGTGGATCAACAGGATGCGGCTTCGCAACGCGTCGCGGT 733

QY 1443 CAAGTTCACGCGGT 1457

Db 734 CAAGTTCACGCGGT 748

RESULT 14

AW424821/c

LOCUS 660039C10.x1 660 - Mixed stages of anther and pollen Zea mays cDNA,

DEFINITION mRNA sequence.

ACCESSION AW424821

VERSION AW424821.1 GI:6952768

KEYWORDS EST.

SOURCE Zea mays

ORGANISM Zea mays

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.

AUTHORS Walbot, V.

TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University

COMMENT Unpublished (1999)

Contact: Walbot V

Department of Biological Sciences

Stanford University

855 California Ave, Palo Alto, CA 94304, USA

Tel: 650 723 2227

Fax: 650 725 8221

Email: walbot@stanford.edu

Plate: 660039 row: C column: 10.

Location/Qualifiers

FEATURES



Mon Mar 8 10:13:27 2004

QY	485	TCAGAGAGTACTCCCTGAAGCTCTCGGTATCTAGTGGCCAGGCATCCAAAGGCGGCAAG	544
Db	481	TCCGGGAGTACTCCCTGAAGCTCTCAGCATCTTGAGCCAGGCTTGCAAGCCGCGCAAG	540
QY	545	TTGTGGACATGCAGGAACCTTACATGAGGATGACGCTGGACTCCATCTGCA	595
Db	541	TTGTGGACATGCAGGAGCTGTACATGAGGATGACGCTGGACTGATCTGCA	591

Search completed: March 6, 2004, 02:18:15  
Job time : 5032 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 6, 2004, 00:53:15 ; Search time 678 Seconds  
(without alignments)  
10281.808 Million cell updates/sec

Title: US-10-021-657-1

Perfect score: 1906

Sequence: 1 gaattggcagcagggaagc.....aaaaaaaaaaactcgag 1906

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 2421054 seqs, 1828716029 residues

Total number of hits satisfying chosen parameters: 4842108

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.\*

1: /cgn2\_6/ptodata/2/pubna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/2/pubna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/2/pubna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/2/pubna/US10A\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/2/pubna/US10B\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/2/pubna/US10C\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/2/pubna/US10\_NEW\_PUB.seq.\*  
17: /cgn2\_6/ptodata/2/pubna/US60\_NEW\_PUB.seq.\*  
18: /cgn2\_6/ptodata/2/pubna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1906	100.0	1906	13	US-10-021-657-1
2	1906	100.0	1906	14	US-10-412-000-1
3	1378.4	72.3	3897	13	US-10-021-657-7
4	1378.4	72.3	3897	14	US-10-412-000-7
5	433.2	22.7	494	13	US-10-021-657-3
6	433.2	22.7	494	14	US-10-412-000-3
7	392.4	20.6	1807	12	US-10-425-114-23429
8	298.2	15.6	1847	12	US-10-425-114-28465
9	269	14.1	1617	15	US-10-260-238-1236
10	255	13.4	1694	12	US-10-425-114-28590
11	228	12.0	1638	15	US-10-260-238-907
12	226.4	11.9	1527	15	US-10-260-238-422
13	217.2	11.4	1680	12	US-10-425-114-1524
14	206	10.8	1806	12	US-10-425-114-28220
15	198	10.4	1761	12	US-10-425-114-26392

16 194.2 10.2 825 12 US-10-425-114-35451  
17 183.8 9.6 1498 12 US-10-425-114-17454  
18 178 9.3 1267 12 US-10-425-114-32933  
19 166.2 8.7 854 15 US-10-260-238-4619  
20 165.4 8.7 1473 9 US-09-938-842A-1981  
21 165.4 8.7 1473 11 US-09-938-842A-1981  
22 163.6 8.6 654 15 US-10-260-238-4618  
23 162.2 8.5 658 15 US-10-260-238-4617  
24 161.6 8.5 444 15 US-10-260-238-5268  
25 160.4 8.4 732 15 US-10-260-238-4608  
26 159 8.3 2431 12 US-10-424-599-120032  
27 157.4 8.3 2239 12 US-10-425-114-10296  
28 149.4 7.8 604 15 US-10-260-238-3255  
29 148.2 7.8 1662 9 US-09-938-842A-881  
30 148.2 7.8 1662 11 US-09-938-842A-881  
31 136.8 7.2 1462 12 US-10-424-599-20189  
32 135.6 7.1 1614 9 US-09-938-842A-2678  
33 135.6 7.1 1614 11 US-09-938-842A-2678  
34 134.8 7.1 2763 12 US-10-424-599-120026  
35 131.8 6.9 1888 15 US-10-429-949-21  
36 129.8 6.8 520 15 US-10-260-238-4618  
37 128.2 6.7 756 12 US-10-424-599-103097  
38 127.6 6.7 620 12 US-10-425-114-4749  
39 120.4 6.3 866 9 US-09-770-445-572  
40 119 6.2 1542 9 US-09-938-842A-1868  
41 119 6.2 1542 11 US-09-938-842A-1868  
42 118.2 6.2 1729 12 US-10-425-114-880  
43 116.8 6.1 1872 12 US-10-424-599-17485  
44 114 6.0 1199 12 US-10-425-114-32508  
45 107.6 5.6 1349 12 US-10-424-599-120029

#### ALIGNMENTS

RESULT 1  
US-10-021-657-1  
; Sequence 1, Application US/10021657  
; Publication No. US00020083483A1  
; GENERAL INFORMATION:  
; APPLICANT: ALBERTSEN, MARC C.  
; APPLICANT: FOX, TIM  
; APPLICANT: HUFFMAN, GARY  
; APPLICANT: TRIMMELL, MARY  
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES MEDIATING MALE FERTILITY AND  
; TITLE OF INVENTION: METHOD OF USING SAME  
; FILE REFERENCE: 1148  
; CURRENT APPLICATION NUMBER: US/10/021.657  
; CURRENT FILING DATE: 2001-12-14  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 1906  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1..1638, 1642..1767)  
US-10-021-657-1

Query Match 100.0%; Score 1906; DB 13; Length 1906;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1906; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GAATTGGCAGCAGGGAAGCTCACTACCGCGGCGAGCCATCGCCATTCCTCCACTA 60  
Db 1 GAATTGGCAGCAGGGAAGCTCACTACCGCGGCGAGCCATCGCCATTCCTCCACTA 60  
QY 61 GCAGGCGCTCAAGTATACGCGCTCTCTGTTGCTCTCATGATCTGTGTCAG 120  
Db 61 GCAGGCGCTCAAGTATACGCGCTCTCTGTTGCTCTCATGATCTGTGTCAG 120  
QY 121 AGGTGGAGCTGAGGAAGCAAGAGCCCGAGATCATGCGCCAGTATCGGCGCAACGGTG 180

121 AGGTGGAGCCTTGAGGAGCAGAAAGGCCCGAGATCATGGCCAGTCACTCGGCGCAACGGT 180  
181 GAGCAGCTGAGGAACTACCAACCGATGACAGCTGGCTTGTGCGGTACTGTCTACGGCAC 240  
181 GAGCAGCTGAGGAACTACCAACCGATGACAGCTGGCTTGTGCGGTACTGTCTACGGCAC 240  
241 AGGACAGTGCAGCTGCAGATGCGGTTCACTTCTACACTATCATCGCTGACCGGTGAAT 300  
241 AGGACAGTGCAGCTGCAGATGCGGTTCACTTCTACACTATCATCGCTGACCGGTGAAT 300  
301 GTCGAGCAGTCTCCTCAAGACTAACCTTCAACCAATACCCCAAGGGAATCGTGTACAGATCC 360  
301 GTCGAGCAGTCTCCTCAAGACTAACCTTCAACCAATACCCCAAGGGAATCGTGTACAGATCC 360  
361 TACATGGACGCTCTCTCGGTGACGGCATCTTCAACCGCGACGGGAGCTGTGGAGGAAG 420  
361 TACATGGACGCTCTCTCGGTGACGGCATCTTCAACCGCGACGGGAGCTGTGGAGGAAG 420  
421 CAGAGGACAGCGGAGTTTCAGTTCGGCTCCAAAGAACCTCAGGGAATTCAGCGCCATT 480  
421 CAGAGGACAGCGGAGTTTCAGTTCGGCTCCAAAGAACCTCAGGGAATTCAGCGCCATT 480  
481 GTGTTTCAGAGAGTACTCCCTGAGCTGTGCGGTATCTAGCGAGGATCCAAAGGAGGC 540  
481 GTGTTTCAGAGAGTACTCCCTGAGCTGTGCGGTATCTAGCGAGGATCCAAAGGAGGC 540  
541 AAAGTTGTGACATGACAGAACTTTACATGAGGATGACCTGAGCTCCATCTGCAAGGTT 600  
541 AAAGTTGTGACATGACAGAACTTTACATGAGGATGACCTGAGCTCCATCTGCAAGGTT 600  
601 GGGTTCGGGTCGAGATCGGACGCTGTGCGGAGATCTCCCGAGAACAGCTTTCGCGCAG 660  
601 GGGTTCGGGTCGAGATCGGACGCTGTGCGGAGATCTCCCGAGAACAGCTTTCGCGCAG 660  
661 GCGTTCGATGCCGCCAACATCATCATCAGCTCGGTTATCATCAACCGCTGTGGCGCATC 720  
661 GCGTTCGATGCCGCCAACATCATCATCAGCTCGGTTATCATCAACCGCTGTGGCGCATC 720  
721 AAGAGTTTTCACGTCGGGTGAGGCGCTCTAGCGCAGAGCATCAAGCTCTGTGGAC 780  
721 AAGAGTTTTCACGTCGGGTGAGGCGCTCTAGCGCAGAGCATCAAGCTCTGTGGAC 780  
781 GAGTTCACCTACAGCGTGATCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 840  
781 GAGTTCACCTACAGCGTGATCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 840  
841 GGCAACAGAGAGAGTGAAGCAGCAGATCTGTTCAGGTTTCATCGAGCTGGGCGAGGCC 900  
841 GGCAACAGAGAGAGTGAAGCAGCAGATCTGTTCAGGTTTCATCGAGCTGGGCGAGGCC 900  
901 GCGGACAGCGCGCGCTTCCGGGAGATGAAGAGCTTCGGGAGCTGGGTCAACTTC 960  
901 GCGGACAGCGCGCGCTTCCGGGAGATGAAGAGCTTCGGGAGCTGGGTCAACTTC 960  
961 GTGATCGCGCGCGCGGACAGCAGCGGACGAGCTGTGTGTTCAGCCCATGGCCATG 1020  
961 GTGATCGCGCGCGCGGACAGCAGCGGACGAGCTGTGTGTTCAGCCCATGGCCATG 1020  
1021 TCCACCGCGGACGTGGCGGAGAGCTGGCGGAGCTGTGTGTTCAGCCCATGGCCATG 1080  
1021 TCCACCGCGGACGTGGCGGAGAGCTGGCGGAGCTGTGTGTTCAGCCCATGGCCATG 1080  
1081 GCGCGGAGGAGGAGCTGACGCTGTCTGCGCGGCGCTGACGCGGAGCAGCAAGGCG 1140  
1081 GCGCGGAGGAGGAGCTGACGCTGTCTGCGCGGCGCTGACGCGGAGCAGCAAGGCG 1140  
1141 TCCGCGCGCGCGGTGGCGAGTTCGCGGCGCTCTCCTACCTACAGAGCTCGGCAAGCTG 1200  
1141 TCCGCGCGCGCGGTGGCGAGTTCGCGGCGCTCTCCTACCTACAGAGCTCGGCAAGCTG 1200  
1201 GTCTACCTCCAGCGCTGCTGACCGAGAGCTGCGGCTGTACCGCGCGCTCTCCTCAGGAC 1260

1201 GTCTACCTCCAGCGCTGCTGTCACGAGACGCTCGGCTGTACCCGCGCTCCTCTCAGGAC 1260  
1261 CCCAAGGGGATCCTGGAGGACGAGCTGCTCCCGGACCGGAGCGAGGTGAGGCGCGCGGG 1320  
1261 CCCAAGGGGATCCTGGAGGACGAGCTGCTCCCGGACCGGAGCGAGGTGAGGCGCGCGGG 1320  
1321 ATGTGACGTACGTGCGCTACTGATGGGCGGATGAGTACAACTCTGGGCGCGGAGCGG 1380  
1321 ATGTGACGTACGTGCGCTACTGATGGGCGGATGAGTACAACTCTGGGCGCGGAGCGG 1380  
1381 GCGAGCTTCGCGCGGAGCGGTGATCAACGAGGATGCGCGGTTCCGCAACGCGTCCGCG 1440  
1381 GCGAGCTTCGCGCGGAGCGGTGATCAACGAGGATGCGCGGTTCCGCAACGCGTCCGCG 1440  
1441 TTCAAGTTACGCGGTTCCAGCGCGGCGCGGAGTCTGCGTGGGCAAGGACTCGGCGTAC 1500  
1441 TTCAAGTTACGCGGTTCCAGCGCGGCGCGGAGTCTGCGTGGGCAAGGACTCGGCGTAC 1500  
1501 CTGCAGATGAAGATGGCGCTGGCCATCTCTCTCGCTTCTACAGCTTCCGCGCTCTGAG 1560  
1501 CTGCAGATGAAGATGGCGCTGGCCATCTCTCTCGCTTCTACAGCTTCCGCGCTCTGAG 1560  
1561 GGGCACCGCGTGCAGTACCGCATGATGACCATCTCTCCATGCGGCGACGCGCTCAAGGTC 1620  
1561 GGGCACCGCGTGCAGTACCGCATGATGACCATCTCTCCATGCGGCGACGCGCTCAAGGTC 1620  
1621 GCGCTCTTAGGCGCGCTGATGTCATGCGGATTTGGATATGGATATGCTCCCGCTTAAT 1680  
1621 GCGCTCTTAGGCGCGCTGATGTCATGCGGATTTGGATATGGATATGCTCCCGCTTAAT 1680  
1681 CCACGACAAATACCGCTGTTTCAAAATTTGATGATGATGATGATGATGATGATGATGATG 1740  
1681 CCACGACAAATACCGCTGTTTCAAAATTTGATGATGATGATGATGATGATGATGATGATG 1740  
1741 GTTTCATTTGGTGGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTT 1800  
1741 GTTTCATTTGGTGGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTT 1800  
1801 TCACAGTGTGTTGATCT 1860  
1801 TCACAGTGTGTTGATCT 1860  
1861 ATATATATATCT 1906  
1861 ATATATATATCT 1906

RESULT 2

US-10-412-000-1  
; Sequence 1, Application US/10412000  
; Publication No. US20030182689A1  
; GENERAL INFORMATION:  
; APPLICANT: ALBERTSEN, MARC C.  
; APPLICANT: FOX, TIM  
; APPLICANT: HUFFMAN, GARY  
; APPLICANT: TRIMMELL, MARY  
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES MEDIATING MALE FERTILITY AND  
; FILE REFERENCE: 1148  
; CURRENT APPLICATION NUMBER: US/10/412,000  
; CURRENT FILING DATE: 2003-04-11  
; PRIOR APPLICATION NUMBER: US/09/670,153  
; PRIOR FILING DATE: 2001-06-11  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 1906  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1..1638, 1642..1767)  
US-10-412-000-1

QY	1021	TCCACCCGGACGTGGCCGAGAACTCGCCGCGAGACTGTGCGCTTCGAGCGGAGCGC	1080
DB	1021	TCCACCCGGACGTGGCCGAGAACTCGCCGCGAGACTGTGCGCTTCGAGCGGAGCGC	1080
QY	1081	GGCGCGAGGAGGGCTCACGCTCGTCTCGCGCGCGCTGACGCCGACGACAAGGCG	1140
DB	1081	GGCGCGAGGAGGGCGTCAACGCTCGTCTCGCGCGCGCTGACGCCGACGACAAGGCG	1140
QY	1141	TTGCGCGCCCGGTGGCGCAGTTTCGCGGCGCTCTCACTACGACAGCCTCGGCAAGCTG	1200
DB	1141	TTGCGCGCCCGGTGGCGCAGTTTCGCGGCGCTCTCACTACGACAGCCTCGGCAAGCTG	1200
QY	1201	GTCTACTCCAAGCTCGGTACCCGAGACGCTCGCCGTATGACCCGCGCTCCCTCAGAC	1260
DB	1201	GTCTACTCCAAGCTCGGTACCCGAGACGCTCGCCGTATGACCCGCGCTCCCTCAGAC	1260
QY	1261	CCCAAGGGGATCCTGGAGGACAGCTGTCTCCGGACGGGACGAAGGTGAGGCGCGGCGG	1320
DB	1261	CCCAAGGGGATCCTGGAGGACAGCTGTCTCCGGACGGGACGAAGGTGAGGCGCGGCGG	1320
QY	1321	ATGGTGACGTACGTGCCCTACTCGATGGGGCGGATGGAGTAACTGGGGCCCCACGCG	1380
DB	1321	ATGGTGACGTACGTGCCCTACTCGATGGGGCGGATGGAGTAACTGGGGCCCCACGCG	1380
QY	1381	GCAGCTTCGGCGCGAGCGGTGGATCAACGAGGATGGCGCGTTCGCAACGCGTCCGCG	1440
DB	1381	GCAGCTTCGGCGCGAGCGGTGGATCAACGAGGATGGCGCGTTCGCAACGCGTCCGCG	1440
QY	1441	TTCAAGTTACGCGCTTCAGGGCGGCGCAGGANTGCTCGGGCAAGGACTCGGCGTAC	1500
DB	1441	TTCAAGTTACGCGCTTCAGGGCGGCGCAGGANTGCTCGGGCAAGGACTCGGCGTAC	1500
QY	1501	CTCGAGATGAAGATGGCGCTGGCCATCTCTTCGCTTACAGCTTCGGGTGCTGGAG	1560
DB	1501	CTCGAGATGAAGATGGCGCTGGCCATCTCTTCGCTTACAGCTTCGGGTGCTGGAG	1560
QY	1561	GGGCAACCGGTGACGATGATGACCATCTCTCCATGGCGCAACGCTTCAAGGTC	1620
DB	1561	GGGCAACCGGTGACGATGATGACCATCTCTCCATGGCGCAACGCTTCAAGGTC	1620
QY	1621	CGCGTCTCTAAGGCGCTGATGTCATGGCGATTTGGATATGATATCGTCCCGCTTAAT	1680
DB	1621	CGCGTCTCTAAGGCGCTGATGTCATGGCGATTTGGATATGATATCGTCCCGCTTAAT	1680
QY	1681	CCACGACAAATAAGCTCGTGTACAATTTGCATGTCATGATGAGGAAAGCATGG	1740
DB	1681	CCACGACAAATAAGCTCGTGTACAATTTGCATGTCATGATGAGGAAAGCATGG	1740
QY	1741	GTCTTCATGGTGGCTTGGCTTAAGCCTTAAAAACTCGTGGGTCTTGGCAACCAACACA	1800
DB	1741	GTCTTCATGGTGGCTTGGCTTAAGCCTTAAAAACTCGTGGGTCTTGGCAACCAACACA	1800
QY	1801	TCACGTAGTGTTCGTACTCTACTCTCAGTGGAGTGTAGTGACAGCATACAAGTTTCATC	1860
DB	1801	TCACGTAGTGTTCGTACTCTACTCTCAGTGGAGTGTAGTGACAGCATACAAGTTTCATC	1860
QY	1861	ATATATATATCCCTTTCTTTAAAAAATAAAAAAAAAAACTCGAG	1906
DB	1861	ATATATATATCCCTTTCTTTAAAAAATAAAAAAAAAAACTCGAG	1906

RESULT 3  
US-10-021-657-7

US-10-021-657-7  
Sequence 7 Application US/10021657

; Sequence 7, Application US/1002

; Publication No. US20

GENERAL INFORMATION:

APPLICANT: ALBERTSEN, MARC

: APPLICANT: FOX, TIM

APPLICANT: FOX, TIM  
APPLICANT: HUFFMAN, GARY

APPLICANT: HUFFMAN, GARY  
APPLICANT: TRIMMEL, MARY

APPLICANT: TRIMNELL, MARY  
TITLE OF INVENTION. NUCLEO

; TITLE OF INVENTION: NUCLEO

; TITLE OF INVENTION: METHOD

FILE REFERENCE: 1148

CONFIDENTIAL





Mon Mar 8 10:13:26 2004

ug-10-021-657-1.rnpb

	D <sub>b</sub>	2528	CTACGACAGCCTCGCAAGCTGGTCTACTCCACGCCTCGCTCACCGAGAAGCTTCGGCT	2587
	Q <sub>y</sub>	1239	GTAACCCCGCGGTCCCT-----	1254
	D <sub>b</sub>	2588	GTACCCCGCGCTCCCTCAGTGAGCGCGCCGACACGGAACCTCGGTCCAGAGCACAGC	2647
	Q <sub>y</sub>	1255	-----CAGGACCCCAGG	1267
	D <sub>b</sub>	2648	ATGCAGTGAGTGAACCTGAAATGCAATGCACATGCATTGCCGCGCGCAGGAGCCCCAAGG	2707
	Q <sub>y</sub>	1268	GGATCTCTGGAGGACGACGTCTCTCCCGACGGGACGAAGTGAAGGCGCGCGGGATGTGGA	1327
	D <sub>b</sub>	2708	GGATCTCTGGAGGACGACGTCTCCCGACGGGACGAAGTGAAGGCGCGCGGGATGTGGA	2767
	Q <sub>y</sub>	1328	CGTAGCTGCCCTACTCGATGGGGCGGATGAGGTACAACCTGGGGCCCCGACGGCGGAGCT	1387
	D <sub>b</sub>	2768	CGTAGCTGCCCTACTCGATGGGGCGGATGAGGTACAACCTGGGGCCCCGACGGCGGAGCT	2827
	Q <sub>y</sub>	1388	TCCGGCGGAGCGGTGATCAAACGAGATGGCCGCTTCCGCAACGGCTCCCGCTTCCAAGT	1447
	D <sub>b</sub>	2828	TCCGGCGGAGCGGTGATCAAACGAGATGGCCGCTTCCGCAACGGCTCCCGCTTCCAAGT	2887
	Q <sub>y</sub>	1448	TCACGGCGTTTCCAGGCGGGCCGAGGANTGCTCTGGGCAAGGACTCGGGCTACCTGCAGA	1507
	D <sub>b</sub>	2888	TCACGGCGTTTCCAGGCGGGCCGAGGANTGCTCTGGGCAAGGACTCGGGCTACCTGCAGA	2947
	Q <sub>y</sub>	1508	TGAAGATGGCGCTGGCCCATCCTCTTCGCTTCTACAGCTTCCGGCTGCTGGAGGGGCACC	1567
	D <sub>b</sub>	2948	TGAAGATGGCGCTGGCCCATCCTCTTCGCTTCTACAGCTTCCGGCTGCTGGAGGGGCACC	3007
	Q <sub>y</sub>	1568	CGGTGCAGTACCGCATGATGACCATCTCTCCATGGCGCACGGCTCAAGGTCGGGTCT	1627
	D <sub>b</sub>	3008	CGGTGCAGTACCGCATGATGACCATCTCTCCATGGCGCACGGCTCAAGGTCGGGTCT	3067
	Q <sub>y</sub>	1628	CTAGGGCGCTCGATGTGTCATGGCGATTGGATATGATATGCTCCCGCTTAAATCCAGCAG	1687
	D <sub>b</sub>	3068	CTAGGGCGCTCGATGTGTCATGGCGATTG-----GGATAATCATCCCGCTTAAATCC-----	3117
	Q <sub>y</sub>	1688	AAATAACGCTCGTGTTACAAATTTTCATGTCATGTCATGTAAGGGAAGCGATGGTTTTCAT	1747
	D <sub>b</sub>	3118	-----TTAAAAATTTGTCATGTCATGTCATGTAAGGGAAGCGATGGTTTTCAT	3163
	Q <sub>y</sub>	1748	TGTTGGCTTGGCTTAAAGCCTTAAAACCTCCGTCGGGTCTTCGGAACCAACCAATCACTAG	1807
	D <sub>b</sub>	3164	TGTTGGCTTGGCTTAAAGCCTTAAAACCTCCGTCGGGTCTTCGGAACCAACCAATCACTAG	3223
	Q <sub>y</sub>	1808	TGTTTTGTACTCTACTCTCAGTGGAGTGTAGTACAGCATACAAGTTCAATCATATATA	1867
	D <sub>b</sub>	3224	TGTTTTGTACTCTACTCTCAGTGGAGTGTAGTACAGCATACAAGTTCAATCATATATA	3283
	Q <sub>y</sub>	1868	TTATCCTCTTTCTTT	1891
	D <sub>b</sub>	3284	TTATCCTCTTTCTTT	3297

```

RESULT 5
US-10-021-657-3
; Sequence 3, Application US/10021657
; Publication No. US20020083483A1
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, MARC C.
; APPLICANT: FOX, TIM
; APPLICANT: HUFFMAN, GARY
; APPLICANT: TRIMMELL, MARY
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES MEDIATING MALE FERTILITY AND
; TITLE OF INVENTION: METHOD OF USING SAME
; FILE REFERENCE: 1148
; CURRENT APPLICATION NUMBER: US/10/021,657
; CURRENT FILING DATE: 2001-12-14
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3

```

```

; LENGTH: 494
; TYPE: DNA
; ORGANISM: Sorghum sp.
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (1)..(494)
; OTHER INFORMATION: "n" bases may be a, t, c, g, other or unknown
US-10-021-657-3

Query Match                22.7%; Score 433.2; DB 13; Length 494;
Best Local Similarity      93.7%; Pred. No. 2.4e-106;
Matches 447; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

Qy      259  ATCCCGTTTCACTTCCTACACCTACATCGCTGACCCCGGTGAATGTCGAGCATGTCTCTCAAG 318
Db      13   ATGCCGTTTCACTTCCTACACCTACATCGCTGACCCCGGTGAATGTCGAGCATGTCTCTCAAG 72

Qy      319  ACTAACTTCACCAATTAACCCCAAGGGAATCGTGTACAGATCCTTACATGGAGCGTCTCTCTC 378
Db      73   ACTAACTTCACCAATTAACCCCAAGGGGAGCGTGTACAGATCCTTACATGGATGTGTCTCTC 132

Qy      379  GGTGAGCGGCATCTTCAACGCGCAGCGGCGAGCTGTGGAGAGAACGAGAGAACGCGGAGT 438
Db      133  GGTGACGGCAATTCAACGCTGACGCGCGAGCTGTGGAGGAAGCAGAGGAACGCGGAGT 192

Qy      439  TTCGAGTTTCGCTTCCAAAGAACCTTACGAGGATTTACGCGGCATGTGTTTCAGAGAGTACTCC 498
Db      193  TTCGAGTTTCGCTTCCAAAGAACCTTACGAGGATTTCAAGTGCATGTGTTTCAGAGAGTACTCC 252

Qy      499  CTGAAGCTGTCCGGTATACTGAGCCAGGCATCCAAAGCCAGGCAAGTTGTGGACATGCAG 558
Db      253  CTGAAGCTGTCCGGGCATACTGAGTCAGGCATCCAAAGCCAGGCAAGTTGTGGACATGCAG 312

Qy      559  GAACTTTACATGAGGATGAGCTGGACTCCATCTTGCAGAGTTGGGTTCCGGGTCGAGATC 618
Db      313  GAACTTTACATGAGGATGACACTGGAATCGATCTGCAANGTTGGTTCCGGGTCGNANATC 372

Qy      619  GGCAAGCTGTCCGAGATCTCTCCCGAGNACAGCTTCGCGCAGGCGGTTTCGATGCCGCCAAC 678
Db      373  GGCAAGCTGTTCNCCGGATCTCTCCCGAGAACAGCTTCNCCCAGGCGGTTTCGATGCCCGCTAAC 432

Qy      679  ATCATCATCAGCTGCGGTTTCATCGACCCCGCTGTGGCGCATCAAGAGGTTCTTCCAC 735

```

```

RESULT 6
US-10-412-000-3
; Sequence 3, Application US/10412000
; Publication No. US20030182689A1
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, MARC C.
; APPLICANT: FOX, TIM
; APPLICANT: HUFFMAN, GARY
; APPLICANT: TRIMNELL, MARY
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES MEDIATING MALE FERTILITY AND
; TITLE OF INVENTION: METHOD OF USING SAME
; FILE REFERENCE: 1148
; CURRENT APPLICATION NUMBER: US/10/412,000
; CURRENT FILING DATE: 2003-04-11
; PRIOR APPLICATION NUMBER: US/09/670,153
; PRIOR FILING DATE: 2001-06-11
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 494
; TYPE: DNA
; ORGANISM: Sorghum sp.
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1)..(494)
; OTHER INFORMATION: "n" bases may be a, t, c, g, other or unknown
US-10-412-000-3

```

	Query Match	20.6%	Score 392.4	DB 12	Length 1807
	Best Local Similarity	58.2%	Pred. No. 3.8e-95		
	Matches	855	Conservative 0	Mismatches 93	Gaps 6
Qy	159	GCCAGTCATCGCGCCCAACGGTGGAGCAGCTGAGGAATACACCGGATGCACGACTGGCT	218		
Db	179	GCCGGTGGTGGGAGCGGTGTTCCACCAAGCTGTACACGTCGCGGGGTGCACGACTACCA	238		
Qy	219	TGTCGGGTACCTGTCAAGGCCACAGGACAGTACCGTGCACATGCGG---TTCACTTCCCTA	275		

QY 1350 GGGATGAGTAACTCGGGCCCGACGCGCGAGCTTCGGCGGAGCGGTGGATCAA 1409  
 Db 1310 CCGGATGAGTAACTCTGGGCAAGGACGCGAGGTTCGGCGGAGCGGTGGCTGA 1369  
 QY 1410 CGAGATGCGCGGTTCGGCAACGCGTGGCGGTTCAGTTACGGGTTCAGCGGGGCC 1469  
 Db 1370 TCAGGACGAGAGTTCAGCAGGAGAGCGGTTCAGTTACAGGCTTCAGGCTGGTCC 1429  
 QY 1470 GAGGATCGCTGGGCAAGGATCGGCGTACCTGGAGATGAAGATGGCGCTGGCCATCT 1529  
 Db 1430 GAGGATCGCTGGGCAAGGATCGGCGTACCTGGAGATGAAGATGGCGCTGGCT 1489  
 QY 1530 CTTCGCTTCTACAGCTTCGCGGTTCGGGCGGCGAC-----CGGTTCAGTACCGAT 1583  
 Db 1490 CTTCCGATCTTCGCTTCGCGCTGGCGAGCGGCAAGCGGTGAATTACCGGAC 1549  
 QY 1584 GATGACCATCTCTCCATCGCGGACGGCT 1613  
 Db 1550 CATGATCAGCTCCACATCGACGAGGCTCT 1579

RESULT 8  
 US-10-425-114-28465  
 ; Sequence 28465, Application US/10425114  
 ; Publication No. US2004003488A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Liu, Jingdong  
 ; APPLICANT: Zhou, Yihua  
 ; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Screen, Steven E.  
 ; APPLICANT: Tabaska, Jack E.  
 ; APPLICANT: Cao, Yongwei  
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
 ; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(5313)B  
 ; CURRENT APPLICATION NUMBER: US/10/425,114  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 73128  
 ; SEQ ID NO 28465  
 ; LENGTH: 1947  
 ; TYPE: DNA  
 ; ORGANISM: Zea mays  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: LIB4757-030-F2\_FLI  
 US-10-425-114-28465

Query Match 15.6%; Score 298.2; DB 12; Length 1947;  
 Best Local Similarity 52.8%; Pred. No. 8.9e-70;  
 Matches 817; Conservative 0; Mismatches 678; Indels 52; Gaps 6;

QY 70 CACAAGTACATCGCGCTCTCTGTTGCTCTCTCATGGATCCTGGTCCAGAGTGGAGC 129  
 Db 65 CCCAGGTGTTCTCCCTGATCTTGTCTCTCATCTAGTGTCTATCTTGGCGCTCCGGCGC 124  
 QY 130 CTGAGGAGCAGAAAGCGCGGAGATCATGGCGAGTATCGGCGCAACGGTGGAGCAGCTG 189  
 Db 125 CGCGCGCGCGGAGAAACCGGAGTACCGCGCGGTGGCTGGCAGCGTCTGCAACAGCTG 184  
 QY 190 AGGACTACACCGGATCAGACTGGCTTGTGGGTACCTGTCTACGCGCAGGACAGTGT 249  
 Db 185 TTCAACTTCGGCGGCTGATGGAGTACCAACCGGCTCTCCGCGAGTACCGTACCTTC 244  
 QY 250 ACCGTGCAATCGGCTTCACTTCTCTACACTACATCGCTGACCGCGGTGAATGTGAGAT 309  
 Db 245 CGCATGCTTACCCGAGCTGAGCTTACATCTACCGTGGAGCGCGCAACGTTGAGTAC 304  
 QY 310 GTCTCTCAAGACTTAACTTCAACAAATTAACCAAGGAAATCGTGTACATCTTACATGAC 369  
 Db 305 ATCTCTCAAGGCAACTTTCGGCAACTACGCGCAAGGAGCAGCGCTGACAGCTGTAGCGAG 364  
 QY 370 GTGCTCTCGGTGAGGCTCTTCAACGCGGAGGAGCTGTGGAGGAGCAGAGGAG 429

Db 365 GACCTCTCGCGGACGCGCATCTTCAACGTGAGCGCGCAAGTGGCGGACACAGCGGAAG 424  
 QY 430 ACGGCGAGTTTCGAGTTTGGCTTCCAGAACTGAGGAGTTTCAGGCGCATTTGTGTTTCA 489  
 Db 425 GTCGCCAGCCAGAGTTTCTCCACGCGGGTGTGCGCGACTACAGCACCGGCGGTGTTCGC 484  
 QY 490 GAGTACTCCCTGAAGCTGTTCGGGTATACGT---AGCCAGGCGATCCAAAGGCAAGGAAAT 546  
 Db 485 GACACGGCGGCGAGCTCGCGGCGATCGTGGCTCCGCGCGCGCGCGCGCGGCGGAGAG 544  
 QY 547 GTGACATGACGAGAACTTTATCAGGATGACGTGGATCCATCTGCAAGGTTGGGTTTC 606  
 Db 545 CTAGACATGACGAGAGTCTGATCGTTCGAGCGGTGGACTCCATCTTCAAGGTCGGGTT 604  
 QY 607 GGGGTGAGATCGGACACGCTCTGCCAGATCTCCCGAGAGACAG---TTCCGCGCAGCG 663  
 Db 605 GGGGTACGCTTGGGCGTCTGTCCGCTCCAGCGAGGAGGCGCGCATTTGCCNAGCG 664  
 QY 664 TTGATGCGCGCAACATCATCATCACGCTGGGGTTTCATCGACCGCTGTGGCGCATCAAG 723  
 Db 665 TTGACAGAGCGCAGCAGCAGGCTGTGCGCGCTTCTTGGACCGCTTCTGGAAGGCCAAG 724  
 QY 724 AGGTTCTTCAACGTGCGGTTCAGAGCCCTCTCTAGCGCAGAGCATCAAGCTCGTGCAGAG 783  
 Db 725 AGGTTCTTCAACGTGCGGTTCAGAGCCCTCTCTAGCGCAGAGCATCAAGCTCGTGCAGAG 784  
 QY 784 TTCACTCAGCGTGTATCGCGGAGGAGGCGGAGATCGTGGAGTTCGGCGCAGCGCGC 843  
 Db 785 TTGCTCTAGCGCGTGTGACAGGAGATCGAGCAGATGAGAGAGAC-----CAACAGGA 840  
 QY 844 AAACAGGAGAGATGAAGCAGCAGCATCTCTCACGGTTTCATCGAGTGGCGGAGCGCGCG 903  
 Db 841 ATTGTGAGCGCCAGAGATATTACCTCTCCGAGTCCGTACATGATGGGTGAGTGCAGAG 900  
 QY 904 GACGCGCGCGCGGTTCGGGAGCAGATGAGAGCTCCCGGAGCTGTGCTGCTTCAACGATG 963  
 Db 901 CGGACGACGCGGATCACGAGAGAGATTAACGTTCTGACACCGGTGACATTTCCCTGACAC 960  
 QY 964 ATCGCGCGCGGAGACACGACGCGCGAGCGCTGTCTGTGTTTCAACGATGCGC---ATGT 1021  
 Db 961 GCCAAGAAAGAGGACATCTCTGTGAGGTTCTTGTGGAGCGGAGAGCGATCCCGGCTGC 1020  
 QY 1022 CCCACCGCGAGCTGGCGGAGAGCTGCGCGCGAGCTGTGCGGTTTCGAGGCGGAGCGCG 1081  
 Db 1021 TCCGACAAAGAGTACCTGCGGAGCATATCTCACTTGTGATCGCGCGCGGACAG 1080  
 QY 1082 CGCGGAGAGGCGCTACGCT-----CGTCTCTGCGCGCGCGCTGACGCGCGAGAC 1134  
 Db 1081 ACCGCGGAGACCGCTGTGCTGTTTCTTACGTGCTCTGAGGAAACAGAGCATCCAGGAC 1140  
 QY 1135 AAG-----GCGTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 1161  
 Db 1141 AGGATCGCGAGAGGTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 1200  
 QY 1162 TTGCGCGCGCTCTCACTACGACAGCTCGGCAAGTGTGTCTACCTCAGCGCTGCGTTC 1221  
 Db 1201 CTGTCGCGGTCTGACCGGAGCGCCATCAGCAAGATGCACTACCTGCAACGCGCGCGCTG 1260  
 QY 1222 ACAGAGAGCTTCGCGCTGTACCGCGCTCTCAGACACCGCGCGCGCGCGCGCGCGCGAG 1281  
 Db 1261 ACAGAGACCGCTTCGCGCTCTACCTTCCGTCGCGCGCGCGCGCGCGCGCGCGCGAG 1320  
 QY 1282 GACGTGTCGCGGACGAGAGGTTGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 1341  
 Db 1321 GACAGTTGCGGAGCGCTAGCTGTCTAAGAGGAGGACATGTTGACTACCGCGCGTAC 1380  
 QY 1342 TCGATGGCGGAGTGGAGTAACTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 1401  
 Db 1381 CAATGGGAGGATGCGGTTCTCTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGAGG 1440  
 QY 1402 TGGATCAAGAGGATGCGGCTTCCGCAACGCTGCGCGCTTCAAGTTTCAAGGCTTCCAG 1461  
 Db 1441 TGGTGGACGACGCGCGCGCTTCTGCTCCCGGAGAGCGCGCTTCAAGTTTCAAGGCTTCCAG 1500

QY	1462	GGGGGGCGAGGATCGCTGGCAGGAGCTCGCGTACCTCGAGATGAAGATGGCGCTG	1521
Db	1501	GGGGGGCCCTCGGCTGCTTGGAAAGGAATTCGGCTACAGGCAGATGAAGATCTTCGCG	1560
QY	1522	GCCATCCTCTTCGGCTTCTACAGCTTCGGCTGCTGGAGGGGCACC	1568
Db	1561	GCTGGCTTCTTACCTGTTTCAGTTTGAGATGGGATGCCAACGC	1607

RESIST 9

```

US-10-260-238-1236
; Sequence 1236, Application US/10260238
; Publication No. US20040016025A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, Paul R.
; APPLICANT: Moughamer, Paul G.
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiya
; APPLICANT: Kreps, Joel
; APPLICANT: P-ovart, Nicholas
; APPLICANT: Ricke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
; FILE REFERENCE: 60111-NP
; CURRENT APPLICATION NUMBER: US/10/260,238
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/325,448
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 6077
; SEQ ID NO 1236
; LENGTH: 1617
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: N_region
; LOCATION: (1326)..(1326)
; OTHER INFORMATION: n = any nucleotide
US-10-260-238-1236

```

Query Match	14.11%	Score	269	DB	15	Length	1617
Best Local Similarity	53.2%	Prod. No.	6.1e-62				
Matches	724	Conservative	0	Mismatches	581	Indels	57
Gaps	5						
QY	282	CATCGCTCACCGGTGAATGTCGAGCATGTCTCTCAAGACTAACTTCACAAATTACCCAA	341				
Db	276	CACCTGCGACCCGGAACAAGTGGCGCACATCTTCAGTCCAACTTCGACAACTACCCAA	335				
QY	342	GGGAATCGTGTACAGATCTTACATGACGTCGTCTCTCGGTGACGCGCATCTTCAACGCCGA	401				
Db	336	GGCGCACGCTGTTCG---CGACATGTTGCACTCTCTCGCGCGCGCATCTTCAACACCGA	392				
QY	402	CGGCGAGCTGTGAGAGAGACAGAGGAAGACGCGGAGTTTCGAGTTTCGCTCTCAAGAACT	461				
Db	393	CGCGGAGCGGTGGCGGCGCACGGAAACAAGGCCAGATGCTCATCACACCGCCGGTT	452				
QY	462	GAGGGAATTCAGCGCCATTTGTTTCAGAGAGTACTCCCTGAAG---CTGTGGGTATTACT	518				
Db	453	CCGCGCTTTCGTGGCGCGTCCAGCTCCACAAAGGTGGAGAAAGGGCTCTCTCCCTTTCCT	512				
QY	519	GAGCCAGGCATCCAGGCGAGGCAAGTTCTGGACATGACGGAACCTTTACATCAGAGATGAC	578				
Db	513	CGCGCACGTCGCGGACGCCCGGAGACGTCGCGACCTTCGAGGACGTTTCACGCGTGGTC	572				
QY	579	GCTGGACTCCATCTGCAAGGTTGGGTTTCGGGTTTCGAGATCGGCACGCTGTCCCAAGTCT	638				



; NUMBER OF SEQ ID NOS: 6077		
; SEQ ID NO 907		
; LENGTH: 1638		
; TYPE: DNA		
; ORGANISM: Oryza sativa		
US-10-260-238-907		
Query Match		
Best Local Similarity 12.0%; Score 228; DB 15; Length 1638;		
Matches 735; Conservative 0; Mismatches 500; Indels 117; Gaps 9;		
QY	290	ACCCGGTGAATGTGCGAGATGCTCTCAAGACTAACTTACCAATTCACCAAGGGAATCG 349
DB	293	ACCCGGGGAACCTTGAGACATCTCGCGCGCGCTTCGAACTACCCCAAGGGCCCCA 352
QY	350	TGTACAGATCTACATGACAGCTGCTCTCGGTGACGGCATCTTCAACGCCGACGGCGAGC 409
DB	353	TGTGGCAGGATCGTTCCAGACCTCTCTCGCCAGGGGCACTTCACTCCGACGGCGAGA 412
QY	410	TGTGGAGGAACGACGAGGAAGACGGCGAGTTTCGAGTTTCGCTCTCAAGAACTGAGGAT 469
DB	413	CGTGGCTGATCCAGCGAAGACGGCAGCCCTCGAGTTTCAACACCGGACGTTGCGCCAGG 472
QY	470	TCAGGCCATGTGTTTCAAGAGTACTCTCTGAA---GCTGTGCGGTATCTAGAGCCAGG 526
DB	473	CCATGGACGCTGGGCCAACCGCTCCATCAAGTACCGCTGTGGGTAATCTTCAGACACC 532
QY	527	CATCCAGGCGAGCAAGTTTGTGACATSCAGGAACCTTATCATGAGGATGACGTGGACT 586
DB	533	ACTGCAAGCGGCCGCAAGCGTGGACCTCCAGGACCTCTCTGCGCTCACTCTCGATA 592
QY	587	COATCTGAAAGTTGGGTTGGGGTTCGAGATCGGCACGCTGTGCGCAGATCTCCCGGAGA 646
DB	593	ACATCTGGGGCTCACTTTCGCAAGAGACCCAGAGACGCTGTCCCGCGCTTGGCGGAGA 652
QY	647	ACAGTTCCGCGAGCGTTCGATGCGGCCAACATCATCATCATCTGCGGTTCATCGACC 706
DB	653	ACCCCTTCGCCACCGCTTTTGACGAGGCCACCGAGCGACGATGACAGGTTCTGTTC 712
QY	707	CGC-----TGTGGCGCATCAAGAGTTCTTCCAGCTCGGFTCAGAGGCCCTCTTAGCGC 760
DB	713	CGAGTTGTTGGCGCATCAAGAGGCTTCGTTGTCGGCAGCGAGGAGCTTCCGCG 772
QY	761	AGAGTCATAGCTGTGACGAGTTCACTACAGCGTGTACGCGGAGGAGGCGGAGA 820
DB	773	ACAGCCTCGCCGTGTGA-----CCGTCACATGACGGAGA 808
QY	821	TGCTCGAGTCCGGCGCAGCGCAACAGGAGAAGATGAAGCAGCATCTCTGTCAACGCT 880
DB	809	CCATCGGCGACG-----CAAGCGACGCCGTCGACGACCTGCTTCCCGGT 856
QY	881	TCATGAGCTGGCGAGCGCGCAGCAGCGCGGGCGGCTTCGGGACGATAGAGCTCC 940
DB	857	TCATGAAGAAGCGCA-----CAGCAAGGCAAGCGTTCCCGGAGGACGTG---CTCC 907
QY	941	GGACGTGTGCTCACTTCTGATCGCGCGGGGACACGACGCGGACGACGCTGCT 1000
DB	908	AGTGGATCGCTCACTTCTGCTCGCGCGGCGCACACGCTGTCGTCGCGCTCAGCT 967
QY	1001	GGTTCCAGCACATGGCCATGTCCACCCCGGACGTGGCGGAGAGCTGCGCCGCGAGCTGT 1060
DB	969	GGTTCTTCTGACTCTCATGACAGAGGGGACGCTGGAGCGCAAGTGGTCTCTGAGAT-- 1025
QY	1061	GCAGCTTCGAGGGGAGCGCGCGCGAGAGGGCGTCAAGCTGCTGCTGCTGCGCGGGG 1120
DB	1026	-----CGCTTCGTTCTCAGGAGACGCGCGGGGACACACACGCGGAGG 1068
QY	1121	CTGACGCCGACGACAGAGCGCTTCCCGCCCGCTGGCGGAGTTGCGGGCTCTCACT 1180
DB	1069	TGAGCG-----GAGGAGCCTCTAACT 1090
QY	1181	ACGACAGCCTCGGCAAGCTGTCTACCTCCAGCGCTTCGTCACCGAGACGCTCCGCGCTGT 1240

DB	1091	TCGACGAGCTGAGCGACTGTGTACTCAAGCGCGCATTCGCGAGACGCTCGCGCTGT 1150
QY	1241	ACCCGCGCTCCTCAGGACCCCAAGGGGATCTGAGAGACGACGCTGCTGCCGACGGGA 1300
DB	1151	ACCCGCTGTGCGCGAGGATTCAGTACGTGTGCGGACGACGCTGCTGCCGACGGCA 1210
QY	1301	CGAAGGTGAGGGCGCGGGGATGGTACGTAGCTGCCCTACTCGATGGGGCGGATGAGT 1360
DB	1211	CAGTGTGCGCGCGCTCGGGATCACCTACTTCCATCTACTCAGTGGGAAGATGAGGA 1270
QY	1361	ACAACTGGGGCCCCGACCGGCGAGCTTCGCGCGGAGCGGTGATCAACGAGATGGCG 1420
DB	1271	GCATATGGGGGAAGGACTGCGCAGAGTTCCGGCGGAGCGGTGCTATCGCGCGACGGCA 1330
QY	1421	CGTTCGCCAA-----CGCTCGCGCTCAAGTTCAAGGCGGGGCGGCGCCAGGA 1474
DB	1331	GCGCTTCGAGCCTGTCAAGGACGCTACCGCTTGTGGGTTCAACGGCGGCGCGGAA 1390
QY	1475	TCGCTCTGGGCAAGGACTCGGCGTACTCTGAGATGAAGATGGCGCTGGCCATCTCTTCC 1534
DB	1391	CGTGGCTCGCAGGACCTGCGCTTACCTGAGATGAATCCATCGCTCGCGCTGCTGTC 1450
QY	1535	GTTTCTACAGCTTCGCGCTCTGAGGGGACCGGTCGAGTACCGCATGATGACCATCC 1594
DB	1451	TGCGCAACTCGGTGAGCTCGTCCGGGCGACAAAGGTGGAGCAGAGATGTGCTGACGC 1510
QY	1595	TCCTCATGGCGCGCGCTCAAGTCCGCGTC 1626
DB	1511	TTTTCATGAAGATGGACTCGGAGTGACGTC 1542

RESULT 12

US-10-260-238-422

; Sequence 422, Application US/10260238

; Publication No. US20040016025A1

; GENERAL INFORMATION:

; APPLICANT: Budworth, Paul R.

; APPLICANT: Moughamer, Todd G.

; APPLICANT: Briggs, Steven P.

; APPLICANT: Cooper, Bret

; APPLICANT: Glazebrook, Jane

; APPLICANT: Goff, Stephen A.

; APPLICANT: Katagiri, Fumiyaki

; APPLICANT: Kreps, Joel

; APPLICANT: Provart, Nicholas

; APPLICANT: Riche, Darrell

; APPLICANT: Zhu, Tong

; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION

; FILE REFERENCE: 60111-NP

; CURRENT APPLICATION NUMBER: US/10/260,238

; CURRENT FILING DATE: 2002-09-26

; PRIOR APPLICATION NUMBER: US 60/325,448

; PRIOR FILING DATE: 2001-09-26

; PRIOR APPLICATION NUMBER: US 60/325,277

; PRIOR FILING DATE: 2001-09-26

; PRIOR APPLICATION NUMBER: US 60/370,620

; PRIOR FILING DATE: 2002-04-04

; NUMBER OF SEQ ID NOS: 6077

; SEQ ID NO 422

; LENGTH: 1527

; TYPE: DNA

; ORGANISM: Oryza sativa

US-10-260-238-422

Query Match

Best Local Similarity 11.9%; Score 226.4; DB 15; Length 1527;

Matches 678; Conservative 0; Mismatches 476; Indels 96; Gaps 7;

QY	266	TCACITCTCTACACCTACATCGTCCCGGTGAATGTCGAGCATGCTCTCAAGACTAAT 325
DB	239	TCACGGCGGGCTCATCCCGCAACCCCGCCATGTCAGTATACGCTGAGACTAAT 298
QY	326	TCACCAATTAACCCCAAGGAATCGTGTACAGATCTCTACATGACGCTGCTCTCGGTGACG 385

Db	299	TTGGTAATACCCAAAGGCGAGCTCGCCGTGTCATGCTCGTCTCGACTTCCTCGGCCATG 358	Db	1286	GGCTAGACGAGCGCGCGCTTCGCGCGGAGAGCCGTTCAAGTACCCGGTTTTCACG 1345
Qy	386	GCATCTCAACGCGCAGCGCAGCTGTGGAGAAAGACAGAGAAAGACGCGAGTTTCAGT 445	Qy	1463	CGGGCCGAGATCTGCTCGGCAAGACTCGCGGTACCTGCAGATGAAG 1512
Db	359	GCATCTTCACTCGACCGCGAGCGAGTGCAGTGGCAGCGAAGCGCCGACGCTACGAGT 418	Db	1346	CGGGCCGAGATGTGCTTCGCGCAAGGAGATGGCTTACATACAGATGAAG 1395
Qy	446	TCGCTCTCAAGAACCTGAGGATTC---AGCGCATTTGTTTCAGAGTACTCCCTGGA 502	RESULT 13		
Db	419	TCAACAAGCGCTCGCTGAGGAATTCGTGGTGGACACCGTCAGGTCGAGGTCTGTCAGA 478	US-10-425-114-1524		
Qy	503	AGCTGTGGGTATCTAGCCAGGATCCAGAGGAGCAAGTGTGACATGACAGGAC 562	; Sequence 1524, Application US/10425114		
Db	479	GGCTCTCGCTGTGAGCGGGGAGCGGAGCGCGGAGCGGCTTCGCTTCGACGAGGA----- 590	; Publication No. US2004003488A1		
Qy	563	TTTACATGAGGATGACGTGGAATCTCATCTGCAAGGTTGGGTCGGGGTCGAGATCGGCA 622	; GENERAL INFORMATION:		
Db	539	TGCTGAGCGCTTCGGTTCGACACATCTGCATGTGCTTCGACGAGGA----- 590	; APPLICANT: Liu, Jingdong		
Qy	623	CGCTGTGCGCAGATCTCCCGAGAACAGCTTCGCGAGCGGTCGATGCGCGCAACATCA 682	; APPLICANT: Zhou, Yihua		
Db	591	----CCGCGCTGCTCGCGGAGGACGATGCG---CTCGCCCGAGAGCGCGAAGTTCA 643	; APPLICANT: Kovalic, David K.		
Qy	683	TCATACGCTGCGGTTTCATCGACCCGCTGTGGCGATCAAGAGTTCTTCCACGTCGGT 742	; APPLICANT: Screen, Steven E.		
Db	644	TGCGCGCTTCACTGACGCGGCAATGCGCGTATGACCGGTTCACTGTCGCGGTCAGT 703	; APPLICANT: Tabaska, Jack E.		
Qy	743	CAGAGCCCTCTAGCGCAGAGCATCAAGCTGTGGACGAGTTCACTACAGCTGATCC 802	; APPLICANT: Cao, Yongwei		
Db	704	CGCGTGGCGCTTCAAGAGCGTTTCAACATGAGCCGAGAGGAGGATGCGGAGGCG 763	; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With		
Qy	803	GCGGAGGAGGCGGAGATGCTGAGGTCCGGGCGAGCGGCAACAGGAGAGATGAAGC 862	; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement		
Db	764	TCGCCACGATCCACGCTTCGCGGAG---CGGATCTGCGGAGCGACGAGGAGAGGGG 820	; FILE REFERENCE: 38-21(53313)B		
Qy	863	ACGACATCTGTACCGTTTCATCGAGCTGGCGAGGCGCGGACGCGCGCGGCTTCG 922	; CURRENT APPLICATION NUMBER: US/10/425,114		
Db	821	AGCCCGGCTGGCGGAGTGACGATCTCTGTGCGCTTCGCGGAGCGGCGGATCA 880	; CURRENT FILING DATE: 2003-04-28		
Qy	923	GGACATAGAGCTCCGCGAGTGTGCTCAACTTCGTGATFCGCGCGGCGGACGACGA 982	; NUMBER OF SEQ ID NOS: 73128		
Db	881	GCACG---AGAGCTCCGTGACGTGGTACCAACTTCCTCATCGCGGCGGCGACGCA 937	; SEQ ID NO 1524		
Qy	983	CGCGCAGCAGCTGTGTGTTTCACGCAATGGCCATGTTCCACCGGAGCTGCGCGAG 1042	; LENGTH: 1680		
Db	938	CGTCAAGCGGTTACCTGTTCTTGTGCTACTGTTCGCGCGCGCGCGAGCGGCGTCA 997	; TYPE: DNA		
Qy	1043	AGCTGCGCGGAGCTGTGCGCTTCGAGCGGAGCGCGCGGAGGAGGCGGTCAAGC 1102	; ORGANISM: Zea mays		
Db	998	AGATCGTACGAGATCCACGCA----- 1020	; FEATURE:		
Qy	1103	TCGTGCTCTCGGCGCGCTGAGCCGACGACAGCGCTTCGCGCGCGCGTGGCGAGT 1162	; OTHER INFORMATION: Clone ID: 700156359_FLI		
Db	1021	--GTGCGCGCGCTTCGCGTGGCACAGCGCACCGACGT----- 1057	US-10-425-114-1524		
Qy	1163	TCGCGGCGCTCTCACTACGACGCTCGGACGCTGTTTACTCCAGCGCTGCTCA 1222	Query Match		
Db	1058	-----TCAACTTGACGAGCTGCGTGATGCAATGCAATGCTCCATCGCGCCATCA 1105	Best Local Similarity		
Qy	1223	CCGAGAGCTCTCGGCTGTACCCCGCGCTCCCTCAGGACCCCAAGGGGATCTGAGGAGC 1282	Matches		
Db	1106	CCGAGTCCATCGGCTGTACCCACCGTGGCCATGAGCTCGACAGCTGCAAGAGGAGC 1165	11.4%; Score 217.2; DB 12; Length 1680;		
Qy	1283	ACGTGCTCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1342	52.2%; Pred No. 5, 4e-48;		
Db	1166	ACTTCTTCCGCGAGCGGAGCTGTGTTGGGAAAGGCTGTTGTTCTTACAGCGGTACG 1225	705; Conservative 0; Mismatches 543; Indels 102; Gaps 6;		
Qy	1343	CGATGGGCGGATGAGTACAACTTGGGCGCGGAGCTTCGCGCGGAGGAGGAGGAGGAG 1402	286 GGTGACCCCGGTGAATGTGAGCATGTCTCAAGATACTTCAACCAATTTACCCCAAGGA 345		
Db	1226	CGATGGCAAGGAGTGTGAGGAGTGTGGGCGCGGAGCTGTGAGGAGTTGAGGCGAGG 1285	342 GCCGACCCCGGCAAGCTGCGCACGTTTCACTGTCCTGTCCTGTCCTGTCCTGTCCTG 401		
Qy	1403	GGATCAAGGAGTGGCGGCTTCGCAACGAGTTCGCGCTTCAAGTTCAAGGCTTCAGG 1462	346 ATCTGTGACAGATCTTACATGAGCATGTCTCTCGGTGAGCGCATCTTCAAGCCGACGCG 405		



Db	1144	-----TCGCGAGATCGCTACGTGCACGCGCCAT	1174
QY	1221	CACGAGACGCTCGGCTGTACCCGCCGTCCTCAGGACCCCAAGGGGATCTGGAGGA	1280
Db	1175	CACGGAGTCGATCGGCTGTACCCGCCGTCGCGGTTAACTCGCTGTCGCGGAGACGGA	1234
QY	1281	CGACGTGCTCCCGGACGGGACGAAGGTGAGGCGCGGGGATGGTACGTACGTGCCCTA	1340
Db	1235	CGACGTCCTGCCCGACGGGACAGCCGTGCTGCAGGGTGTTCTGGCGGTACAATTCATA	1294
QY	1341	CTCGATGGGGCGGATGAGTACAACTGGGGCCCGACGCGCGGAGCTTCGCGCCGGAGCG	1400
Db	1295	TGGGATGGGGCGGATGAGTCSGTGTGGGGCGACGACGCGCGGGAATACCGCGCGGAGCG	1354
QY	1401	GTGGATCAA---CGAGGATGGCGGTTCCCGAACGCGTCGCCGTTCAAGTTACAGCGGTT	1457
Db	1355	GTGGCTGAACCCCAAGACGGAACGTTCCGCGCGGACAGCCCCCTTCGGTTCTGGTGGCATT	1414
QY	1458	CGAGCGGGGCGGAGGATCTSCCTGGGCAAGGACTCGGCGGTACTCGAGATGAAGATGGC	1517
Db	1415	TCAGCGGGGCGGAGGCTTGTCTGGGCAAGGAATGGCGTACATCCAGATGAAGTCCAT	1474
QY	1518	GCTGGCCATCCTCTTCGCTTCTACAGCTTCGGGCTGCTGGAGGGGCAACCGGTGCAAGTA	1577
Db	1475	CGTGGCGGTGCTGGAGGAGCTCGAGTGGCGGTTGAOCGGCGCTACCGGCCCGCGCA	1534
QY	1578	CCGCATGATGACCAATCCTCTCCATGGGCGACGGGCTCAAGTCCGCGT	1625
Db	1535	GGTTGCGTTCGTAACGCTGCGGATGGGAGCGGCTACCTGTGACCGT	1582

```

RESULT 15
US-10-425-114-26392
; Sequence 26392, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 26392
; LENGTH: 1761
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LrB4171-038-F6_FLI
US-10-425-114-26392

```

355	QY	AGATCTTACATGGAGCTGCTCTCTCGGTACAGGCATCTTCAACGGCCGACGGCGAGCTGTGG	414
420	Db	GCGGCGCTGCTCGGGACCTCTCGGGCGGGCATCTTCAACGTGGACGGCCACGCCTGG	479
415	QY	AGGAAGCAGAGGAGACGGCGAGTTTCAGATTCTCGCTTCCAGAAACCTTGAGGATTC---	471
480	Db	CGCCACAGCGCAGATGGCCAGCCTCGAGCTCGGCAGCGTCCGCGTCCGCTCCTACGCG	539
472	QY	AGCGCAATGTGTTCCAGAGGATCTCCCTGAAGCTGTGCGGTATCTAGGACGAGCATCC	531
540	Db	TTCCGGATCGTGGCGCAGAGGTGGAGTCCCGCTCTGTCGCGGTCTCGCGCGCGCGCC	599
532	QY	AAGGACGGAAGTTGTGCACATCAGAACTTTATCATGAGATGACCTGATCTCCATC	591
600	Db	GACGCTGGCGGCTCGTGACCTGCAGACGTTTCGGCGCTTCGCGGCTTCGACCAATC	659
592	QY	TGCAAGTTGGTTCGGGGTCCAGATCCGCACTGTGCGCAGATCTCCCGAGAACAGC	651
660	Db	TGCAAGATCTCTTCGGCTGACCCCGGATGCTTGAGCGGGAGATGCCGTGTGAGG	719
652	QY	TTTCGGCAGCGGTTCCATGCCGCCAACATCATCATCACGCTCGGTTTCATCGACCG---	708
720	Db	CTCGCGCAGCGTTCCGACTGCCCAAGCGGCTGTGCGCATTCGGCGGGCGCGCGCTCG	779
709	QY	-----CTGTGGCGATCAAGAGGTTCTTCAAGTCGGGTGAGAGCCCTCTAGCGCAG	762
780	Db	CCGCTAGTGTGAGGGCCAAAGCGCCTGCTGAACGTTCGGTCCGAGAGGAGTCCGGGAG	839
763	QY	AGCATCAAGCTCGTGACAGTTCACTACAGCTGATCCGCGCGGAGGAAGCCGAGATC	822
840	Db	GCCATCGCGCTCGTGCAGAGCTCGGGCGCGCTGATCCGGAGCGCCGGAAGCTGGGA	899
823	QY	GTGAGGTTCCGGGCCAGCGGCAAAACAGGAGAGATGAAGCAAGACATCTGTCAAGTTTC	882
900	Db	GTCTGTCGGCAG-----CCAGCACTCTGTCTCCGCTTC	932
883	QY	ATCGAGCTGGCGAGCCCGGCGACGACGGCGGCTTCGGGAGCGATGAAGACCTCCGG	942
933	Db	ATGCGCTTCGGCGCGCGCCCGGCA-----CGGCGACGAAAGTACCTCCGC	977
943	QY	GAGTGTTGCTCAACTTCGTGATTCGCGCGCGGGAACACGACGGCGACGCTGTGTGG	1002
978	Db	GACATGTTGTGAGCTTCTGTGGCGGGCGGACACCGTGTCTCCGCGCTGAC-----	1033
1003	QY	TTCAAGCAGATGCGCCATGTCCACCCCGGAGCTGGCCGAGAAAGTGGCGCGAGCTGTGC	1062
1034	Db	-----CAGCTGTTC	1043
1063	QY	GCGTTGAGCGGAGCGCGCGCGGAGGGCGTCACGCTCGTGTCTCGCGCGCGCT	1122
1044	Db	ATGCTCTGTCTCCAGAAACCGGGCGTGGCGCGCCATCGCGCG-----	1088
1123	QY	GACGCCAGACAAGGGGTTTCGCGCGCGCGTGGCGAGTTTCGCGGCTCTCTCACTAC	1182
1089	Db	-----GAGGCGGCGACGGCGAGTCCCGCGCGCTCACTAC	1127
1183	QY	GACAGCTCGGCAAGTGTCTACCTCCACGCTCGGTACCGAGAGCTTCGCGCTGTAC	1242
1128	Db	GAGCACTCAAGGGCTTGACTACCCACGCGGTGTGTACGAGAACATGCGGCTGTTC	1187
1243	QY	CCGCGCTCGCTCAGGACCCCAAGGGATCTCTGGAGCAGCTGTTCGCGACGGGACG	1302
1188	Db	CCGCGGTGAGATTTCGACTCAAGATTTCGCGGGGCGGACGCTGTCCCGGACGGCACC	1247
1303	QY	AAGTGAAGGCGCGCGGATGTGAGTACGTGCTTCCCTACTTCGATGGGGGATGAGATAC	1362
1248	Db	TACGTGCGCGGGGACACGCGGTGAGCTTACACCCCTACGCAATGGCGGATGCCGCG	1307
1363	QY	AACGTGGGCGCCGACGCGGGGAGTTTCGGCGCGGAGCGGTGGATCAACGAGGATGGCG--	1420
1308	Db	ATCTGGGGCGGACACAGCGGCTTCGCGCGGACCGGTGGTCTACCGGAGCGGCGCC	1367
1421	QY	-CGTTTCGCAACGGGTCCGCTTCAAGTTCACGGGCTTCAGGGGGGCGAGGATCTGC	1479

Db	1368	TCGTTTCCTCCAGAGAGCCTGTATAGGTACCCGGTGTTCAGGCGGGCCCTCCGCGTCTGC	1427
QY	1480	CTGGGCAAGGACTTCGGCGTACCTGCAGATGAAGATGGCGCTGGCCATCCTCTTCCGCTTC	1539
Db	1428	CTCGGCAAGGAGCTCGCCGTCACCGAGATGAAGTGGTGGTGGCCCGTGGGTGAGGGCG	1487
QY	1540	TACAGCTTCGGGCTGCTGGAGG	1561
Db	1488	TTCGAGTCGAGGTGGTCCGGG	1509

Search completed: March 6, 2004, 04:20:59  
Job time : 707 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 5, 2004, 16:59:14 ; Search time 782 Seconds  
(without alignments)  
10354.304 Million cell updates/sec

Title: US-10-021-657-1  
Perfect score: 1906  
Sequence: 1 gaattggcagcgggaagc.....aaaaaaaaaaaaaacctcgag 1906

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_29Jan04:\*  
1: Geneseq1980s:\*  
2: Geneseq1990s:\*  
3: Geneseq2000s:\*  
4: Geneseq2001as:\*  
5: Geneseq2001bs:\*  
6: Geneseq2002s:\*  
7: Geneseq2003as:\*  
8: Geneseq2003bs:\*  
9: Geneseq2003cs:\*  
10: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1906	100.0	1906	6	ABL60611 Maize SBM
2	1378.4	72.3	3897	6	ABL60615 Maize SBM
3	433.2	22.7	494	6	ABL60612 Sorghum S
4	271.6	14.2	1626	2	Aax26878 Codon mod
5	271.6	14.2	2181	2	Aax26874 Cytochrom
6	226.4	11.9	1527	7	ADA70596
7	214.4	11.2	1563	7	ADA70663
8	212.8	11.2	1536	7	ADA71342
9	165.4	8.7	1473	6	ABZ14176
10	159	8.3	1506	7	ADA70631
11	150.8	7.9	1395	8	ADA48303
12	148.2	7.8	1662	6	ABZ13076
13	144.6	7.6	657	8	ACL21310
14	137.2	7.2	621	8	ACL21306
15	135.6	7.1	1614	6	ABZ14873
16	135.6	7.1	1896	3	ADA45033
17	130.8	6.9	528	8	ADA49310
18	120.4	6.3	866	6	ABN98804
19	119	6.2	1542	6	ABZ14063
20	119	6.2	1736	2	AAX58407
21	108.8	5.7	843	8	ADA48919
22	102.4	5.4	399	8	ACL21305
23	102.4	5.4	460	8	ACL21309 DNA clone

C	24	102.4	5.4	507	8	ACL21308
C	25	102.4	5.4	527	8	ACL21304
	26	102.2	5.4	1365	7	ADA71125
	27	98.8	5.2	1694	3	AAC37317
	28	97	5.1	1816	3	AAC39955
C	29	97	5.1	95223	3	AAF22282
	30	95.6	5.0	1117	3	AAC40695
	31	95.2	5.0	1533	3	AAC42704
	32	94	4.9	1440	3	AAC42927
	33	92	4.8	1488	6	ABZ13809
	34	92	4.8	1488	7	ADA68342
	35	91.2	4.8	1407	3	AAC33115
	36	90	4.7	1605	7	ADA69913
	37	90	4.7	2000	7	ADA71986
	38	87.4	4.6	1512	3	AAC42674
	39	87.2	4.6	1714	3	AAC40480
	40	86.8	4.6	589	3	AAC38695
	41	86.8	4.6	1711	3	AAC51594
	42	86.6	4.5	1584	7	ADA69866
	43	85.4	4.5	1551	3	AAC43122
C	44	84.6	4.4	665	6	ABQ5871
	45	83	4.4	1828	2	AAX58402

ALIGNMENTS

RESULT 1  
ABL60611  
ID ABL60611 standard; cDNA; 1906 BP.  
XX ABL60611;  
AC ABL60611;  
XX 27-AUG-2002 (first entry)  
XX Maize SBMu200 protein encoding cDNA.  
XX SBMu200; gene; fertility; maize; hybrid seed; sterility; plant;  
XX transgenic; ss.  
XX Zea mays.  
XX Key  
FH Location/Qualifiers  
FT 1. 1770  
FT /tag= c  
FT /product= "fused SBMu200 protein"  
FT /partial  
FT /note= "CDS encodes the N-terminal and C-terminal portion  
of SBMu200, separated by a stop codon. CDS lacks an  
initiation codon"  
FT 1.1641  
FT CDS  
FT /tag= a  
FT /product= "N-terminal portion of SBMu200"  
FT /partial  
FT /note= "CDS lacks an initiation codon"  
FT 1642.1770  
FT CDS  
FT /tag= b  
FT /product= "C-terminal portion of SBMu200"  
FT /partial  
FT /note= "CDS lacks an initiation codon"  
XX WC00226789-A2.  
XX 04-APR-2002.  
XX 25-SEP-2001; 2001WO-US029886.  
XX 26-SEP-2000; 2000US-00670153.  
XX (PION-) PIONEER HI-BRED INT INC.  
XX Albertsen MC, Fox T, Huffman G, Trimmell M;

[illegible]





QY	513	TATACCTGAGCCAGGATCCAGGAGGCAAAAGTTGTGGACATGTCAGGAACATTTACATGAG	572
DB	483	CATCTTGGCCGACGCGGCGCAGGCGAGGAGGATCTCCAGGACCTCTCTCTCCG	542
QY	573	GATGACGCTGAGACTCCATCTGCAAGGTTGGGTTCGGGTCGAGATCGGACGCTGTGCGC	632
DB	543	CCTCACCTTCGACAAACATCTGGGCGCTGGCTTCGGCAAGGACCCGAGAGCGCTCGCCCA	602
QY	633	AGATCTCCCGAGAACAGCTTCGGCGAGGGGTCGATGCGGCCAACATCATCATCACGCT	692
DB	603	GGGCTTCGCGGAGACGAGTTGCTCCCGGTTTCGACCGCGCCACCCGAGGCCACGCTCAA	662
QY	693	GGGTTTCATCGACCCG-----CTGTGGGCGCATCAAGAGGTTCTTCCACGTTCGGGTCA	746
DB	663	CCGCTTCACTTCCTCCGAGGTTCTGTGGCGCTGCAAAAAGTGGCTGGCGCTCGCATGA	722
QY	747	GGCCCTCTAGCGCAGAGCATCAAGTCTGTGGAGGATTCACCTACAGCGTGTGTCGCGC	806
DB	723	GACACGCTGACCGAGCATGCGCCACGTCGACCGGCTACCTCGCGCCGCTCATCAAGAA	782
QY	807	GAGGAGGCGAGATCGTCGAGTTCGGGCGCAGCGCAACAGAGAGATGAAGACGA	866
DB	783	GCGCAGCTCGAGCTCGCGCGCGGCGCAACGGAATATGCGACACGCGCGCAGCAGCA	842
QY	867	CATCTCTCACGGTTTCATCGAGCTGGCGGAGCGCGCGCAGCGCGCGCTTCGGGGA	926
DB	843	CTGCTCTCCGGTTTCATCGGAAGGTTCTTACTCGGAG-----	883
QY	927	CGATAAGAGCTTCGGGACGCTGTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG	986
DB	884	-----AGTGGCTCCAGCAGCTGGCGCTCAATCTTCTCTCGCGCGCGCAGCCTCTC	938
QY	987	GACGACGCTGTGTTGTTTCAGGCACATGGCCATCTCCACCCGCGACCTGGCGGAGAGCT	1046
DB	939	CGTGGCGCTTCCTGGTTCTTCTGGCTGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCG	998
QY	1047	GCGCGCGAGCTGTGCGGTTTCGAGGCGGAGCGCGCGCGCGCGCGCGCTCACGCTGT	1106
DB	999	CGTGGCGAGCTGCTC-----CGT	1019
QY	1107	GCTTCGCGCGCGCTGACGCGCAGCAAGCGGTTCCGCCCGCGCGCTGGCGAGTTCGC	1166
DB	1020	TCTCGCGCGTCACGGGCGGCCATGACCGGCAIT-----GTGGCTGGC	1064
QY	1167	GGGCTCTCTACCTACGACAGCTCGGCAAGCTGTCTACCTCCAGCGCTCGGTCCACGA	1226
DB	1065	GGAGCCCTTACCTTCGAGGAGCTCGACCGCTGTCTTCTCAAGGCGCGCTGTGGA	1124
QY	1227	GACGCTCGCGCTGTACCCCGCGCTCCCTCAGGACCCCAAGGGGATCTTGGAGGAGAGCT	1286
DB	1125	GACCTTCGCGCTTACCCCTCGTCCCGAGGAGCTCCAAAGCAGCTGCTCGGCGAGCACTA	1184
QY	1287	GTCGCGGAGCGGACGAAGGTCGAGCGCGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT	1346
DB	1185	CCTCCCGCAGGACCTTCTGTCGCGCGCGGTCGTCGCTTCTCTACTCTCATATCTCGC	1244
QY	1347	GGGCGGATGAGTAACTTGGGCGCCGACGCGCGGAGCTTCCCGCGCGAGCTTCCCGCGGAGCGGTGAT	1406
DB	1245	GGGCGGATGAGGAGGTCGTCGCGCGGAGGAGTGTGTCGAGTTCGCGCGCGGAGGATGGCT	1304
QY	1407	CAACGAGGATGGCG---GTTCCGCAAGCGTCGCGCTTCAAGTTTCAAGTTTCAAGCTTCCAGC	1463
DB	1305	GTCGCGCAGCGCACCAAGTTTCGAGCAGCACACTCGTACAAAGTTTCTGGGCTTCAACGC	1364
QY	1464	GGGCGGAGGATCTGCTGGGCAAGGATCTGCGGAGTCTGTCGAGATGAGATGAGATGGCTGCG	1523
DB	1365	CGGCGGAGGTCGCTGGGCAAGGACCTTACGCTTCTGTCGAGATGAGAGATCATGCGCG	1424
QY	1524	CATCTCTTCGCTTCTACAGCTTTCGCGGTCGTCGAGGCGGCGCGCGCTGAGTACCGAT	1583
DB	1425	GAGCGTGTCTCGGCAACGCTGACCTGGCGCGCGCGCGCGCGCTGAGCAGAGAT	1484

RESULT 4  
AA26878  
ID AAX26878 standard; cDNA; 1626 BP.  
XX  
AC  
AA26878;  
XX  
23-JUN-1999 (first entry)  
XX  
DE Codon modified cDNA encoding a cytochrome P450.  
XX  
KW Cytochrome P450; protein expression; yeast; mutant; ss.  
XX  
OS Synthetic.  
OS *Triticum aestivum*.  
XX  
Key Location/Qualifiers  
FT 1..1626  
FT CDS /\*tag= a  
XX  
FR2768748-A1.  
XX  
PD 26-MAR-1999.  
XX  
PF 24-SEP-1997; 97FR-00012094.  
XX  
PR 24-SEP-1997; 97FR-00012094.  
XX  
PA (RHON ) RHONE-POULENC AGROCHIMIE.  
XX  
PI Batard Y, Schalk M, Durst F, Merck RD;  
XX  
DR WPI: 1999-217499/19.  
DR P-P5DB; AAY01648.  
XX  
PT DNA for expression in yeasts - with codon changes based on yeast codon usage.  
XX  
PS Example 4; Page 24-26; 31pp; French.  
XX  
CC The present sequence encodes a modified cytochrome P450, which was created by altering the codons of AAX26874 to exemplify the invention.  
CC The specification describes a DNA sequence that encodes a protein of interest and contains regions with a high content of codons poorly suited to yeasts, where a sufficient number of these codons are replaced by corresponding codons suited to yeasts in these regions. Yeasts transformed with vectors such DNA sequences can be cultured to produce the protein of interest (especially a plant protein) or, when the protein is an enzyme (such as cytochrome P450), can be cultured in the presence of a substrate for the enzyme to produce a substrate conversion product  
XX  
SQ Sequence 1626 BP; 250 A; 583 C; 514 G; 279 T; 0 U; 0 Other;  
  
Query Match 14.2%; Score 271.6; DB 2; Length 1626;  
Best Local Similarity 54.5%; Pred. No. 3.3e-42;  
Matches 736; Conservative 0; Mismatches 524; Indels 90; Gaps 6;  
  
QY 276 CACCTACATCGTACCGGTCGAATGTCGAGCATGTCCTCAAGACTTAACCTACCAATTA 335  
DB 243 CACCGTCACTGCGACCCGCGCAACTGTCGAGCAGCTCTGAAGCGCGCTTCGACAACTA 302  
QY 336 CCCCAAGGGAATCGTGTACAGATCCCTACATGAGCGTGTCTCTCGGTGACGCGATCTTCAA 395  
DB 303 CCCCAAGGCGCCCTTCTGGCAGCGCGCTTCGCGGACCTCTCGGCGAGCGGCACTTCTCAA 362  
QY 396 CGCGGAGCGGAGCTGTGGAGGAGAGAGAGAGAGAGAGTTCGAGTTCGCGCTC--- 452  
DB 363 TTCGAGCGGCGACACCTTGGCTCGCGCAGCGCAAGAGCGCGCGCTCGAGTTCACCCCG 422  
QY 453 CAGAGAACTGAGGATTTACGCGCATGTGTTCAGAGAGTACTTCCCTGAAGCTGTGGG 512  
DB 423 CAGCTCCGAGCGGCAATGTCCCGCTGGGTCTCGCGCTCCATCCACGGCGCGCTCTGCC 482

```

QY 1584 GATGACCATCCTCCATCGCGCACGGCCT 1613
Db 1495 GTGCTCAGCCTTTCATGAAGGGGGCT 1514

RESULT 5
AA226874
ID AAX226874 standard; cDNA; 2181 BP.
AC AAX226874;
XX
XX
XX
XX 23-JUN-1999 (first entry)
DE Cytochrome P450 cDNA designated CYP86A5.
XX
XX Cytochrome P450; protein expression; yeast; ss.
XX
XX Triticum aestivum.
XX
XX Location/Qualifiers
FH Key 112..1737
FT CDS /*tag= a
FT
XX
XX FR2768748-Al.
XX
XX 26-MAR-1999.
XX
XX 24-SEP-1997; 97FR-00012094.
XX
XX 24-SEP-1997; 97FR-00012094.
XX
XX (RHON ) RHONE-POULENC AGROCHIMIE.
XX
XX Batard Y, Schalk M, Durst F, Werck RD;
XX
XX WPI; 1999-217499/19.
DR P-PSDB; AA01647.
XX
XX DNA for expression in yeasts - with codon changes based on yeast codon
XX usage.
XX
XX Claim 20; Page 21-23; 31pp; French.
XX
XX The present sequence encodes a cytochrome P450, and was used to exemplify
CC the invention. The specification describes a DNA sequence that encodes a
CC protein of interest and contains regions with a high content of codons
CC poorly suited to yeasts, where a sufficient number of these codons are
CC replaced by corresponding codons suited to yeasts in these regions.
CC Yeasts transformed with vectors such DNA sequences can be cultured to
CC produce the protein of interest (especially a plant protein) or, when the
CC protein is an enzyme (such as cytochrome P450), can be cultured in the
CC presence of a substrate for the enzyme to produce a substrate conversion
CC product
XX
XX Sequence 2181 BP; 395 A; 722 C; 684 G; 380 T; 0 U; 0 Other;
XX
XX Query Match 14.2%; Score 271.6; DB 2; Length 2181;
XX Best Local Similarity 54.5%; Pred. No. 3.5e-42;
XX Matches 736; Conservative 0; Mismatches 524; Indels 90; Gaps 6;

QY 276 CACCTACATCGTGTGACCGGTTGAATGTCGAGCATCTCTCAGACTAACTTACCAATTA 335
Db 354 CACCTGTACCTCGACCGCGCAACCTGGAGCACGTCCTTGAAGCGCGCTTCGACACTA 413
QY 336 CCCCAAGGAATCGTGTACAGATCCTACATGACGTCGTCTCTCGGTGACGGCATCTTCAA 395
Db 414 CCCCAAGGCCCCCTTCTGCAAGCGGTCTTCCGGACCTGTCTCGGAGCGCATCTTCAA 473
QY 396 CGCCGACGCGAGCTGTGTGAGGAAGCAGAGAGAGCGGATTTTCGAGTTCCGCTC--- 452
Db 474 TTCCGACGCGACACCTGTCTCGCGCAGCGCAAGACGGCGCGCTCGAGTTTCAACCCCG 533
QY 453 CAAGAACCTGAGGGATTTTCAGCGCCATTTGTTTCAGAGAGTACTCCCTGAAGCTGTGCGG 512

```

Db 1536 GAGCGTGTGCTCCGGCACCGCCTGACCGTGGCGCGGCCACCAGGTGAGCAGAGAT 1595  
QY 1584 GATGACCACTCTCTCATGCGCGACGGCCT 1613  
Db 1596 GTCGCTCAGCTCTTCATGAAGGCGGGCT 1625  
RESULT 6  
ID ADA70596 standard; DNA; 1527 BP.  
AC ADA70596;  
XX 20-NOV-2003 (first entry)  
XX Rice gene, SEQ ID 3919.  
XX Plant; bacterial infection; fungal infection; viral infection; rice;  
KW gene; ds.  
XX Oryza sativa.  
XX WO2003000898-A1.  
XX 03-JAN-2003.  
XX 22-JUN-2001; 2001WO-IB001105.  
XX 22-JUN-2001; 2001WO-IB001105.  
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;  
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;  
XX WPI; 2003-175296/17.  
XX Identifying at least one gene involved in plant resistance or response to  
PT pathogenic infection for conferring resistance or tolerance to a plant to  
PT bacterial, fungal or viral infection by determining or detecting plant  
PT gene expression.  
XX Claim 6; SEQ ID NO 3919; 999pp; English.  
XX The present invention relates to a method (M1) for identifying genes  
CC involved in plant resistance or response to pathogenic infection. M1  
CC comprises identifying a gene whose expression is significantly altered in  
CC the incompatible interaction of plant gene expression relative to  
CC expression of the gene in an uninfected plant, in a mutant plant that  
CC does not express a gene associated with response to pathogenic infection,  
CC or in a corresponding incompatible or compatible interaction. (M1) is  
CC useful for conferring resistance to resistance or tolerance to a plant to  
CC bacterial, fungal or viral infection. The present sequence was used to  
CC illustrate the invention.  
XX Sequence 1527 BP; 280 A; 473 C; 494 G; 280 T; 0 U; 0 Other;  
SQ Query Match 11.9%; Score 226.4; DB 7; Length 1527;  
Best Local Similarity 54.2%; Pred. No. 1.3e-33;  
Matches 678; Conservative 0; Mismatches 476; Indels 96; Gaps 7;  
QY 266 TCATCTTCCATACCTACATCGCTGACCGCGTGAATGTCGAGCATGCTCTCAAGATCACT 325  
Db 239 TCACCGCGCGCTCATACCGGCAACCGCGGCAATGTCGAGTATACGCTGAAGATCACT 298  
QY 326 TCACCAATTAACCCAGGGAATCGTGTACAGATCTCTACATGACGCTCTCTCGGTGAG 385  
Db 299 TTGGTAATACCCCAAGGCGAGCTCGCGTGTCCATGTCTGCTGCTCTCTCGGCGCATG 358  
QY 386 GCATCTTCAACCGCGCGAGCTCTGGAGGAAGCAGAGGAGAGCGGCGAGTTTCAGT 445  
Db 359 GCATCTTCAACTCCGACGGGAGCACTGGCAGTGGCAGCGGAAGGCGCGCAGCTACGAT 418

QY 446 TCGCTCTCAAGAACCTGAGGATTTTC---AGCGCAATTTGTGTTTCAGAGAGTACTCCCTGA 502  
Db 419 TCAACAAGCGCTCGTGGAGAACTTCGTGGTGACACCGTCAGGTCCGAGGTCTGTCGAGA 478  
QY 503 AGCTGTGGGTATATGAGCCAGGCAATCAAGGCGAGCAAGTTTGTGACATGCGAGAAC 562  
Db 479 GGCTGTGCTGCTGTGAGCGGCGAGCGCGCGAGCGATGCG---CTCGCCCAAGAGCGGAATTCA 538  
QY 563 TTTATACAGGATGACGCTGGACTCCATCTGCAAGGTTGGGTTTCGGGGTTCAGATCGGCA 622  
Db 539 TGCTGGAGCGCTTCGGGTTTCGACAACTGTCATGTGCTTCGACAGGA-----590  
QY 623 CGCTGTGCGCCAGATCTCCCGGAGAAAGCTTCGCGAGGCGCTTCGATGCGCGCAATCA 682  
Db 591 ----CCCGCGTGCCTCGCCGAGGACAGCATGCG---CTCGCCCAAGAGCGGAATTCA 643  
QY 683 TCATACAGCTGGGTTTCATCGACCGCTGTGGGATCAAGAGGTTCTTCCACGTCGGT 742  
Db 644 TGGCGCGCTTCAGTGAACGCGAGAAATCCGCTGATGACCGGTTTCATGTGCGCGTCAAGT 703  
QY 743 CAGAGCCCTCTAGCGCAGAGCATCAAGCTCGTGACGAGTTCACCTACAGCGTGATCC 802  
Db 704 CGCGTGGCGCTTCAAGAGCTGTTTCAATGAGCCCGAGGCGAGATCGGAGGCGCG 763  
QY 803 GCGGAGGAAGCCGAGATCGTGAAGTCCGGCCAGCGGCAAAACAGGAGAGATGAAGC 862  
Db 764 TCGCCACGATCCACGCTTCGCGAG---CGGATCGTCCGGAGCGCAGGAGAGGGG 820  
QY 863 ACACATCTCTGTCACGCTTCATCGAGCTGGCGAGCGCGCGAGCGCGCGGCTTCG 922  
Db 821 AGGCGCGCTGGCGCGCAGTGACGACTTCTGTCGCGCTTCGCGCGAGCGCGGATCACA 880  
QY 923 GGCAGCATAGAGCTCCGCGAGCTGCTCAACTTCGTGATCGCGCGCGGAGCACGA 982  
Db 881 GCGAGC---AGAGCTTCCGTCAGCTGTTGACCACTTCCTCATCGCGCGCGAGACGA 937  
QY 983 CGGCGACGACGCTGTGCTGTTTCAACGACATGCGCATGTCCACCGCGAGCGTGGCCGAGA 1042  
Db 938 CGTCGAGCGGCTTGACCTGCTTCTTCTGCTACTGTCCGCGCGCGCGCGAGGAGCA 997  
QY 1043 AGTTCGCGCGAGCTGTGCGGTTTCAGGCGGAGCGCGCGCGAGGAGGCGGTCAAGC 1102  
Db 998 AGATCGTACGCGAGATCCACGCA-----1020  
QY 1103 TCGTGTCTCGCGCGCGCTGAGCGCGACGACAGGCGTTTCGCGCGCGCGGTGGCGCAGT 1162  
Db 1021 --GTGCGCGCGCTTCGCTGGGCAAGGCGCGCGCGCGT-----1057  
QY 1163 TCGCGGCGCTCTCTCACTACGACAGCTTCGCGAAGTGGTCTACTCGACGCTTCGCTCA 1222  
Db 1058 -----TCAACTTGGACGAGCTGGTGTGACATGCAATACCTCATGCGCGCGCATCA 1105  
QY 1223 CCGAGACGCTCCGCGCTGACCGCGCGCTCCCTCAGGACCGCGCGCGAGGATCTCGAGGAGC 1282  
Db 1106 CCGAGTCCATGCGGCTGTACCCACCGCTGGCCATGAGCTCGACAGCTGCAAGGAGGAGC 1165  
QY 1283 ACGTGTCTCGCGACGCGAAGGCTGAGCGCGCGCGGATGTCAGCTGTCGCTTACT 1342  
Db 1166 ACTTCTTCGCGAGCGGACGCTTCGTGGGAAAGGCTGTGTTGTCTTACAGCGGTACG 1225  
QY 1343 CGATGGGCGGATGAGTACAACTGGGCGCGCGAGCGCGAGCTTCGCGCGGAGCGGT 1402  
Db 1286 CGATGGCAGCGGTGAGGACATCTGGGCGCGGAGTGTGAGGAGTTTACGGCGAGCGGT 1285  
QY 1403 GGATCAAGAGAGTGGCGGCTTCGCGACGCTCGCGCTTCAAGTTTCAAGTTTCCAGG 1462  
Db 1286 GGTAGACGAGCGCGCGCTTCGCGCGGAGAGCGCGCTTCAAGTTTCCAGGTTTCCAG 1345  
QY 1463 CGGCGCGGAGGATGCTGCTGGCGAAGGACTCGCGCTTACTTCAGATGAAG 1512  
Db 1346 CGGGCGCGAGGATGCTGCTTCGCGAAGGAGTGGCTTACATACAGATGAAG 1395

	RESULT 7	
	ADA70663	
	ADA70663 standard; DNA; 1563 BP.	
XX	ADA70663;	
XX	20-NOV-2003 (first entry)	
XX	Rice gene, SEQ ID 3986.	
DE	Plant; bacterial infection; fungal infection; viral infection; rice;	
XX	gene; ds.	
OS	Oryza sativa.	
PN	WO200300898-A1.	
PD	03-JAN-2003.	
PF	22-JUN-2001; 2001WO-IB001105.	
PR	22-JUN-2001; 2001WO-IB001105.	
PA	(SYGN ) SYNGENTA PARTICIPATIONS AG.	
PI	Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;	
PI	Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;	
XX	WPI; 2003-175290/17.	
DR	Identifying at least one gene involved in plant resistance or response to	
PT	pathogenic infection for conferring resistance or tolerance to a plant to	
PT	bacterial, fungal or viral infection by determining or detecting plant	
PT	gene expression.	
PS	Claim 6; SEQ ID NO 3986; 899pp; English.	
XX	The present invention relates to a method (M1) for identifying genes	
CC	involved in plant resistance or response to pathogenic infection. M1	
CC	comprises identifying a gene whose expression is significantly altered in	
CC	the incompatible interaction of plant gene expression relative to	
CC	expression of the gene in an uninfected plant, in a mutant plant that	
CC	does not express a gene associated with response to pathogenic infection,	
CC	or in a corresponding incompatible or compatible interaction. (M1) is	
CC	useful for conferring resistance to resistance or tolerance to a plant to	
CC	bacterial, fungal or viral infection. The present sequence was used to	
XX	illustrate the invention.	
SQ	Sequence 1563 BP; 221 A; 572 C; 530 G; 240 T; 0 U; 0 Other;	
	Query Match 11.2%; Score 214.4; DB 7; Length 1563;	
	Best Local Similarity 52.5%; Pred. No. 2.4e-31;	
	Matches 738; Conservative 0; Mismatches 536; Indels 132; Gaps 7;	
QY	175 ACGGTGAGCAGCTGAGGAACCTACCACCGGATGCACACTGGTCTGTTCGGGTACTGTCA 234	
DB	169 ACGGGTGCTGGCGGAGGAGTTCCACACCTCGCGGACTGGTAGCGCACCTGTCTCGC 228	
QY	235 CGGCACAGGACAGTACCGGTGACATGCGGTTCATTCTTCACTACCTAATGCTGACCCG 294	
DB	229 CGCTCGCGGACGGAAACCGGTCCACGTCCACGTCTCGGTGACCGTCACTCGGCGAACCCG 288	
QY	295 GTGNAATGTCAGCATGTCCTCRAGACTAATTCACCAATTAACCAAGGGAATCGTGTAC 354	
DB	289 GCGAACTGTCAGACAATGCTCAGGACGGGTTCGACAACTTCCCAGGGGAGGCCCTTC 348	
QY	355 AGATCTTACATGACGTGCTCTCCGTGAGCGGCATTTCAAACCGCCAGCGGAGCTGTGG 414	
DB	349 GCCCGCTCTCTCGCGACTCTCTCGGCGAGCGCATCTTCAACGTGACGCGCACGCGTGG 408	
QY	415 AGGAAGCAGAGGAAGACGGCGAGTTTCGAGTTCGCTCTCAAGAACCTGAGGGAATTCAGC 474	
DB	409 CGCCACCAAGGAGATGGCCAGCTCGAGCTCGGAGCGTGCCTCGCGCTCTCTACGCC 468	



513 TATACTGAGCCAGGCAATCAAGCCAGGCAAAAGTTGTGGACATGACAGCAACTTTACATGAG 572  
414 TTTTATTTGGGAGTTTGTCTCTCTCGAAATCTTTTGTATTTTCAGATATGTTGATGAA 473  
573 GATGACGCTGGACTCCATCTCGAAAGTTGGGTTCCGGGTCCGATCCGACCGCTGTGCGCC 632  
474 ATGTACATTGGACTCCATCTTTAAAGTTGGATTCGGTGTGGAGTTAGGATGCTCTAGATGG 533  
633 AGATCTCCCGA---GACAGCTTCCGACGCGCTTCGATCCGCGACACATCATCATCAC 689  
534 GTTTAGCAAAAGAGGGGAAGATTCATGAAGGCTTTTGATGAAGCAACCGCGCAACTAG 593  
690 GCTGCGGTTTCATCGACCCGCTGTGCGCATCAAGAGGTTCTTCCACGTCGGGTCAGAGGC 749  
594 TTCACGGGTACCGACCCGTTTGAAGCTGAAATGTTTTCTTAACTTGGATCAGAGTC 653  
750 CCTCCTAGCGCAGAGCATCAAGCTCGTGAACGAGTTTCACTACAGCTGTATCCCGCGAG 809  
654 AAGACTCAAGAGAGCATTTGATTATAGACAAGTTTGTCTATAGTCTCATTTACCACTAA 713  
810 GAAGCCGAGATCGTTCGAGGTCGGGCCAGCGCAACAGGAGAAGATGAAGCAGACAT 869  
714 AAGAAAGAGCTTCCAAGGAACAGAAACA-----CTTCTGTTAGAGGACAT 761  
870 CTTGTCACGCTTCATCGAGCTGGCGCAGGCGCGCAGACGCGCGGCTTCGGGACGA 929  
762 CCTATCGAATTTCTTCGAGAGTCAGAAAGATCCGGAGA-----ACATGAATGA 812  
930 TAAGAGCCTCCGGACGTTGTGCTCAACTCTGTGATCGCGCGGGGACACGACGGCGAC 989  
813 TAAGTACTCGAGGATATCATTTTGAATGTATGTTGTCTGTGTAGGACCAACCGCTGTC 872  
990 GACGCTCTGTGTTTCACGACATGCGCATGTCCCAACCGGACGTGGCGGAGAAAGCTGG 1049  
873 ATCACTCTCTTGTTCTTTGATCTCTGCAAAACCCACATTGTTTCAGAGAA-----926  
1050 CCGGAGCTGTGCGCTTCGAGCGGAGCGCGCGCGGAGGAGGCGTCACGCTCGTGCT 1109  
927 -----AATCGTTCA 935  
1110 CTGCGGCGGCTGACGCCGACAGAGCGTTTCGCGCGCGGCTGGCGCAGTTCCGGG 1169  
936 AGAAATCAGAGATGTGACATCAAGTCAAGAGAAAACAACGATGTAAATGGTTTCATTGA 995  
1170 CTTCTCACTACACAGCCTTCGCAAGCTGTGTACCTCCAGCCTGGGTGCTCCAGAC 1229  
996 AAGTGTACCGAGAAGCTCTTCTCAGATGCAATCTCCATCGCGCCTTGTCTGAGAC 1055  
1230 GTCGCGCTGTACCCCGGCTCCCTCAGGACCCCAAGGGATCTCGAGAGCACACGTGCT 1289  
1056 TATGAGGCTTACCCACCTGTCCTGAGCACAATGAGTGTGCGAATAATGATGACGTTCT 1115  
1290 GCGGACGGGACGAAGGTGAGGGCGCGCGGATGGTGAAGTGCCTTACTCTGATGGG 1349  
1116 TCCAGATGACATAGATGAGCAAGGGGATAATCTACTACATATCTTATGCAATGGG 1175  
1350 GCGGATGGAGTACAACTGGGGCCCGACGCGCGAGCTTCGGCGGAGCGGTGGATCAA 1409  
1176 TAGGATGACTTACATTTGGGGTCAAGATGCTGAGGAAATCAAGCGAGAGATGGCTCA- 1234  
1410 CGAGGATGGCGGTTCCGCAACGCTGCGCTTCAAGTTACGCGGTTCCAGGCGGGGCC 1469  
1235 --AGGACGGGATTTCCAAACCCGAATCAAAATTCAAATTCATAAGCTTCATGCTGGTCC 1292  
1470 GAGGATCTGCCTGGGCAAGGACTCGGCGTACTCGAGATGAAGATGGCGCTGGCCATCCT 1529  
1293 AAGAACTGTAATGGCAAGGATTCGACATCCGCGAGATGAAGATAGTATCAATGGCACT 1352  
1530 CTTCCGCTCTACAGCTCCGCGCTGCTGGAGGGGCAC 1566  
1353 TCTTCACTCTTTTCGCTTCAAAATGGCTGATGAGAAC 1389

QY	473	GCGCCATTGTTTACAGAGTACTCCCTGAAGCTGTCGGGTATACAGCCAGGATCCCA	532
Db	446	ACAAATCGTCGCCAGGAGTGGAGCCCGCTCATGCCGTGCTGCCAACGCCGCGCG	505
QY	533	AGGCAGGCAAGTTGTGGACATGCAAGAACTTTACATGAGGATGACGCTGGACTCCATCT	592
Db	506	ACAGCGCGCGGTGGTCGACCTGCGAGGAGCTGTTCCGCCGCTTCGCTTCGACACCATCT	565
QY	593	GCAAGTTGGGTTTCGGGGTCGAGATCGGCACGCTGTGCCAGATCTCCCGAGAACAGCT	652
Db	566	GCAAGATCTCTTCGGCTTCGACCCGCGCTGCTCGACCGGAGATGCCGCTGTCGGAGC	625
QY	653	TCGCGCAGGGGTTTCGATGCCGCCAACATCATCATCGCTCGGTTTCATCGACCC	707
Db	626	TCGCGCAGCGGTTTCGACCGCGCTGCGCGCTGTCGCGCATGCTGCGGCGCGCGTCCG	685
QY	708	-----GCTGTGGCGCATCAAGAGTTCTTCCAGTGGGTTCAGAGGCCCTCTCTAGCGCAGA	763
Db	686	CGTTGCTGTGAAGATGAAGCGTTTCTCAACGTCGGTTCGAGAGGAGCTCAAGAAG	745
QY	764	GCATCAAGTCTGTCGAGAGTTTCACTACAGCTGATTCGCGGAGGAGGCCGAGATCG	823
Db	746	CCATCAAGCTCATCGACGGCTTCGCGCGCGGATGATCCGGGAGCGCGGAAGCTTGGCG	805
QY	824	TCGAGGTCGGGCGCCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	883
Db	806	TCG-----CGAACAGCCACGACCTCTCTTCCCGGTTCA	838
QY	884	TCGAGCTGGCGAGGCGCGGCGGCGGCGGCGGCTTCGGGAGCATGAAGAGCCTCCGGG	943
Db	839	-----TGCGCTCTTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	878
QY	944	ACGTGCTGCTCAACTTTCGTGATTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	1003
Db	879	-----CAACTTCTCTCTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	920
QY	1004	TCACGCATATGCCCATGTCCACCCGCGAGCTGGCGGAGAGCTGGCGGCGGAGCTGTGCG	1063
Db	921	-----CACCACCTCTGTTC	934
QY	1064	CGTTTCGAGCGGAGCGCGGCGGCGGCGGAGGAGGCTCACGCTGCTGCTGCGCGCGCGGTG	1123
Db	935	TGATCTCTGTCCAAAGAACCCCGAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	994
QY	1124	AGCGCGAGCAAGGCGTTTCGCGCGCGCGGCTGGCGAGTTCGCGGCGCTCTCACTACG	1183
Db	995	CGCGGCGAGAGC-----GCCGCGGTTCAGTACG	1021
QY	1184	ACAGCTTCGGCAAGCTGTCTACCTCCACGCTCGGTTCACCGAGAGCTTCGCGCTGTACC	1243
Db	1022	AGCACCTGAGCGGCTGAATACACCCACGCGGTGTGTACGAGAACATCGCGGTGTTC	1081
QY	1244	CGCGCGTTCCTTCAGGACCCCAAGGGGATCTCGAGGAGCGAGCTGTCGCGGAGCGGAGCA	1303
Db	1082	CGCGGTGCGAGTTTCGATCCAAAGTTCTTGGCGCGCGCGCGAGCTGTCTCCCGAGCGGACCT	1141
QY	1304	AGGTGAGGCGCGGCGGAGTGTGAGTACGTGCGCTACTCGATGGGCGCGATGGAGTACA	1363
Db	1142	ACGTGAGCGGCGGCGCGCGGTATGTACACACCTTCACCCATGGCGCGCATGCCGCGCA	1201
QY	1364	ACTGGGCGCGGAGCTTCGCGCGGCGGAGCTTCGCGCGGAGCGGTGG-----ATCAACGAGGATGGCG	1420
Db	1202	TCTGGGCGCGGAGCTGCGACCGGTTCGCGCGGAGCGGTGGCTTAACCGGCGCGCGCGGCG	1261
QY	1421	CGTTTCGCGCAACGCGTTCGCGGTTCAAGTTCAAGGTTTCAGGCGGCGCGGAGATCTGCC	1480
Db	1262	CGTTTCGTCGGAGAGCGCTCTTCAAGTACCCCGGTTTCCAGGCGCGGCGCTCCGCGTGTGC	1321
QY	1481	TGGGCAAGAGCTCGGCGTACTCTGAGATGAAGATGGCGTGGCCATCTCTTCGCGCTTCT	1540
Db	1322	TCGGCAAGGAGCTCGCCATCACCGAGATGAAGCGGTGAGCGTTCGCGCTGTCGAGGAGCAT	1381

RESULT 10  
ADA70631  
ID  
ADA70631 standard; DNA; 1506 BP.  
XX  
ADA70631;  
XX  
20-NOV-2003 (first entry)  
DT  
Rice gene, SEQ ID 3954.  
DE  
Plant; bacterial infection; fungal infection; viral infection; rice;  
KW  
gene; ds.  
OS  
Oryza sativa.  
XX  
WO2003000898-A1.  
PN  
03-JAN-2003.  
PD  
22-JUN-2001; 2001WO-IB001105.  
XX  
22-JUN-2001; 2001WO-IB001105.  
PR  
(SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX  
Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;  
PI  
Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;  
XX  
WPI; 2003-175290/17.  
DR  
XX  
Identifying at least one gene involved in plant resistance or response to  
PT  
pathogenic infection for conferring resistance or tolerance to a plant to  
PT  
bacterial, fungal or viral infection by determining or detecting plant  
PT  
gene expression.  
XX  
Claim 6; SEQ ID NO 3954; 899pp; English.  
PS  
XX  
The present invention relates to a method (M1) for identifying genes  
CC  
involved in plant resistance or response to pathogenic infection. M1  
CC  
comprises identifying a gene whose expression is significantly altered in  
CC  
the incompatible interaction of plant gene expression relative to  
CC  
expression of the gene in an uninfected plant, in a mutant plant that  
CC  
does not express a gene associated with response to pathogenic infection,  
CC  
or in a corresponding incompatible or compatible interaction. (M1) is  
CC  
useful for conferring resistance to resistance or tolerance to a plant to  
CC  
bacterial, fungal or viral infection. The present sequence was used to  
CC  
illustrate the invention.  
XX  
Sequence 1506 BP; 208 A; 540 C; 521 G; 237 T; 0 U; 0 Other;  
SQ  
Query Match 8.3%; Score 159; DB 7; Length 1506;  
Best Local Similarity 51.0%; Pred. No. 7.9e-21;  
Matches 716; Conservative 0; Mismatches 530; Indels 159; Gaps 8;  
QY 176 CGTGGACAGCTGAGGAACTACACCGGATGCACGACTGCGTGTGCGGTTACCTGTCAC 235  
Db 146 CGGGTCTGTGAGGAGGAGTTGCCAACCTCGCGGACTGTGATGCGGCACTGCTCGGCC 205  
QY 236 GGCACAGACAGTACCGTGCACATCGCGTTCACTTCTCATACCTATACATGCTGACCCGG 295  
Db 206 GCTCGCGGAGCGGACCGTTCACGTCACGTCCTCGGTGTCACCGTCAACGCGGAGCCGG 265  
QY 296 TGAATGTGAGCATGCTCTCAAGACTAACTTCAACAAATTAACCAAGGAAATCGTGTACA 355  
Db 266 CGAAGTGGAGTACATGCTCAAGACGCGCTTCGACAACTTCCCCAAGGGAGGCGGTTGCG 325  
QY 356 GATCTCATATGACATGCTCTCGGTGAGCGCATTTCAAGCCGACCGGAGCTGTGGA 415  
Db 326 CGCGCTCTCTCGCGACCTCTCTCGCGAGCGGCATCTTCAAGCTTCGACGCGGAGCGGTGGC 385  
QY 416 GGAAGCAGAGGAGCGGCGGAGTTTCGAGTTCGCTCCCAAGAACCTGAGGGATTTC---A 472  
Db 386 GCCACAGCGGAGATGGCCAGCCTCGAGCTGGGAGCGGTGCGCGTGAGATCTTACGCGT 445

QY	1541	ACAGCTTCGGGCTGCTGAGGGGCA	1565	QY	416	GGAAAGCAGAGGAGACGCGGAGTTTCGAGTTCGCCTCCAGAACTGAGGGAATTCAGCG	475
Db	1382	TCGACGTCGAGGTCGTCGGCGAGAA	1406	Db	386	GCCACACGCGGAGATGGCCAGCCTCGAGCTGGGAGAGGTGCGGTGAGATCCTACGCGT	445
RESULT 11				QY	476	CCATTGTGTTTCAGAGAGTACTCCCTGAA--GCTGTGCGGGTATATCTAGACGAGGATCCA	532
ID	ADA48303	standard; DNA; 1395 BP.		Db	446	ACAAGATCGTCCGACGAGGTGGAGGCCCGCTCATGCCGTGCTCGCCAAACGCGCG	505
AC	ADA48303;			QY	533	AGGACGCAAAAGTTGTGACATGACAGAACTTTACATGAGATGACGCTGACTTCATCT	592
XX	20-NOV-2003	(first entry)		Db	506	ACAGCGCGCGGTGTCGACCTGACGAGTGTTCGCGCGTTCGCTTCGACCACTCT	565
DE	Rice gene conferring disease resistance in plants.			QY	593	GCAGGTTGGTTCGGGTTCGAGATCGGCACGCTGTGCGCAGATCTCCCGAGAACACT	652
XX	disease resistance; pathogen tolerance; plant pathogen; ds; gene; plant.			Db	566	GCAGATCTCCTTCGCGCTCGACCGGCTGCTCGGAGATGCGCGGATGCGCGGAGC	625
XX	Oryza sativa.			QY	653	TGCGCAGGCGTTCGATCGCGCAACATCATCATCACTGCTGCGGTTCTATCGACCC	707
PN	WO2003000906-A2.			Db	626	TGCGCGACGCGTTCGACCGCGCTCGCGCTGTCCGCTATGCTGCGCGCGCGCTCGC	685
XX	03-JAN-2003.			QY	708	-----GCTGTGCGCATCAAGAGTTCTTCCAGCTCGGTTCCAGAGCCCTCTTAGCGCAGA	763
XX	21-JUN-2002; 2002WO-1B002453.			Db	686	CCTTCTGTGGAAGATGAAGCGGTTTCTCAAGCTCGGTCGAGAGGAGCTCAAGAAGG	745
XX	22-JUN-2001; 2001US-0300112P.			QY	764	GCATCAAGCTCGTGACGAGTTCACTACAGCGTTCATCGCGGAGGAGGAGGAGGAGTCA	823
PR	26-SEP-2001; 2001US-0352277P.			Db	746	CCATCAAGCTCATGACGCGGCTCGCGCGCATGATCC-----GGGAGCGCGGAGC	799
PR	22-MAR-2002; 2002US-0366535P.			QY	824	TCGAGTTCGCGCGCGGCAACAGAGAGAGATGAAGACGACATCTCTGTCACTGTTCA	883
XX	(SYGN ) SYNGENTA PARTICIPATIONS AG.			Db	800	TTGGCGTCGGAACAGCACGACCTCTCTGCGGTTTCATGCGCATCTGCTGAACTTCC	859
XX	Glazebrook J, Briggs S, Cooper B, Goff SA, Moughamer T;			QY	884	TCGAGTTCGCGCGCGGCGGAGCGGCGGCTTCGCGGAGGAGGAGGAGGAGGAGGAGG	943
PI	Katagiri F, Xreps J, Provart N, Ricke D, Zhu T;			Db	860	TCTCTCGCGCGCGGAGCACGCTGCTCTCGCGCTCACCACTCTGTTTCATGA---TCTCT	916
XX	WPI; 2003-184052/18.			QY	944	AGTGTGCTCAACTTCGTGATCGCGCGGCGGAGACAGACGCGGAGGAGGAGGAGGAGG	1003
DR	P-PSDB; ADA48304.			Db	917	CMAGAACCCGACGCTGGCGCGGCTATGCGCGGAGGCGCGCGCGCGCGCGGAGGAG	976
XX	New polynucleotide comprising a plant nucleotide sequence having an open			QY	1004	TCAGGCACATGGCCATGTCCACCCCGGAGCTGCGCGGAGGAGGAGGAGGAGGAGGAGG	1063
PT	reading frame that encodes a polypeptide associated with disease			Db	977	GCSCCGCGCTCAGCTACGAGACCTGAAGCGGCTGAACCTACACCCAGCGCTCTTACG	1036
PT	resistance, useful for conferring resistance or tolerance to a plant			QY	1064	CGTTCAGGCGGAGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1123
PT	pathogen.			Db	1037	AGAACATGCGGCTGTTCCCGCGCGGTGAGTTTCACTTCAAGTTCTGGCGCGCGCGGAGC	1096
XX	Claim 1; SEQ ID NO 373; 299pp; English.			QY	1124	ACGCCAGACAAAGGCGTTTCGCGCGCGGCTGCGCGAGTTTCGCGGCGGCTCTCTCACTACG	1183
XX	The invention relates to a novel isolated polynucleotide comprising a			Db	1097	TGCTCCCGACGCGACCTACGTCGAGCGGCGCGCGGCTGATGATGATGATGATGATGATG	1156
XX	plant nucleotide sequence having an open reading frame that encodes a			QY	1184	ACAGCTTCGCAAGTGTGTTTACCTCCACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1223
CC	polypeptide associated with disease resistance or its fragment having			Db	1157	TGGCGCGATGCGCGGATCTGCGGCGCGCGGCGCGGCGCGGCGCGGCGCGGCGGCGG	1196
CC	substantially the same activity as the full-length polypeptide. The			RESULT 12			
CC	polynucleotide of the invention is useful for conferring resistance or			ID	ABZ13076	standard; DNA; 1662 BP.	
CC	tolerance to a plant pathogen. The present sequence represents a gene			XX	ABZ13076		
CC	conferring disease resistance used in the invention.			AC	ABZ13076;		
XX	Sequence 1395 BP; 196 A; 497 C; 475 G; 227 T; 0 U; 0 Other;			XX	21-JAN-2003	(first entry)	
XX	Query Match 7.9%; Score 150.8; DB 8; Length 1395;			DT	Arabidopsis thaliana	stress regulated gene SEQ ID NO 881.	
XX	Best Local Similarity 49.2%; Pred. No. 2.8e-19;			DE	Arabidopsis thaliana;	plant; gene; stress; transgenic; ds.	
XX	Matches 522; Conservative 0; Mismatches 517; Indels 21; Gaps 4;			XX	Arabidopsis thaliana.		
QY	176	CGGTGGACGCTGAGGAATACCAACGCGATGACGAGTGTGCGGTGACCTGTGCTAC	235	XX	WO200216655-A2.		
Db	146	CGGGTCTGGAGGAGGTTCGCCAACCTCGGCGATGTGTACGCGGACCTGTCTCGCC	205				
QY	236	GGCAGGACGATGACCGTTCGATGCGGTTCACTTCTTACCTACCTACCTACCTACCTAC	295				
Db	206	GCTCCCGAGGGCCGCTCCAGTCCAGTCTCTCGGCTGACCTGACCTGACCTGACCTGAC	265				
QY	296	TGAATGTGAGATGCTTCAGACTTACCAATTTACCCAGGGAATCGGTACA	355				
Db	266	CGAAGCTGGAGTACATGTCTCAAGACGCGCTTCGCAACTTCCCAAGGGAGGCGGCTCG	325				
QY	356	GATCTTACATGACGCTGCTCTCGGTGAGCGGATCTTCAACGCGGATCTTCAACGCGGAG	415				
Db	326	CGCGGCTCTCTCGGCGACCTCTCTCGGCGAGGATCTTCAACGCGGAGCGGCGGCGG	385				

XX	28-FEB-2002.	Db	724	GTCAGCTTGAGCCGGAGCCTAGGAGAGATCGATGGTATTAGATGCTGCTTAATACA	783
PD					
XX	24-AUG-2001; 2001WO-US026685.	Qy	808	AGGAAGCCGAGATCGTCGAGTCCGGCCAGCGGCAACAGGAGAAAGATGAAGCAGC	867
XX	24-AUG-2000; 2000US-0227866P.	Db	784	CGTAAGCAAGAAATGCTGAGTCAGCAGAGAGTGG-----GGTCCAGCGTCAGCGAT	837
PR	26-JAN-2001; 2001US-0264647P.	Qy	868	ATCTGTACACGGTTTCATCAGACTGGGGAGCCCGGACGACGCGCGGGTTCGGGGAC	927
PR	22-JUN-2001; 2001US-0300111P.	Db	838	CTCCTCTCTGTTTCATGAAGAAGA-----CCAGTCGTACAGC	879
XX	(SCRI ) SCRIPPS RES INST.				
PA	(SYGN ) SYNGENTA PARTICIPATIONS AG.	Qy	928	GATAAGAGCCTCCGGGACGTGCTCACTTGTGATCGCGGGCGGACACGACGGCG	987
XX	Harper JF, Kreps J, Wang X, Zhu T;	Db	880	GAGAGCTTCTTACGACACGTGGCGCTTAACCTCATCTAGCTGGACGTGACACGTCATCA	939
XX	WPI; 2002-304127/34.	Qy	988	ACGAGCTGTCTGCTTCAACGACATGCCATGTCCACCCCGGACGCTGCCGAGAAGCTG	1047
DR		Db	940	GTAGCGTTGAGCTGTTTTTCTGGCTCATCAGACGATCCGACGTTGAGGATAAGATC	999
XX	Identifying a stress condition to which a plant cell has been exposed and	Qy	1048	CGCCGCGAGCTGTCGCGCTTCGAGCGGAGCGCGCGGAGGGGCTCAAGTCTGCTG	1107
PT	producing plants with increased tolerance to these abiotic stresses.	Db	1000	GTCCGGGAGATGCTC-----CGTCTGATTGACACACGTGGAAC	1040
PS	Claim 144; SEQ ID NO 881; 577pp + Sequence Listing; English.	Qy	1108	CTCTCGCGCGGCTGACGCCGACACAGAGGCTTCGCGCGCGCTGCGCGAGTTCGCG	1167
CC	The invention relates to identifying a stress condition to which a plant	Db	1041	CGATGTATCTGCTGCTGAGC-----GCG	1062
CC	cell has been exposed, comprising: (a) contacting nucleic acid	Qy	1168	GGCTCTCTCCTACCTAGACAGCCTCGSAAAGCTGCTACTCCACGCCCTCGCTCACCGAG	1227
CC	representative of expressed polynucleotides in the plant cell with an	Db	1063	GAGCGCTTGGAAATTCATGAGTGCACCGCTGTTGTTTACTTGAAGCCGCGCTCTCTGAG	1122
CC	array or probes representative of the plant cell genome; and (b)	Qy	1228	ACGCTCGGCTGTACCCCGCTCCCTCAGGACCCCAAGGGGATCTCTGAGGACGACGTG	1287
CC	detecting a profile of expressed polynucleotides in the plant cell	Db	1123	ACGTTGAGGCTTTACCGCTGCTGCGGGAAGATTCAAAGCACGCTGCTGAACGAGATC	1182
CC	characteristic of a stress response. The method is useful in the	Qy	1288	CTCCCGGAGCGGACGAAGTGTAGGCGCGGCGGATGCTGACGTACGTGCCCTACTCGATG	1347
CC	production of transgenic plants, cells and seeds and in producing plants	Db	1183	TTACCGGACGGAACCTTTCGTACCGCGGGATCGCTCGTGACTTATTTCGATCTACGCGCG	1242
CC	with increased tolerance to abiotic stress. The present sequence is that	Qy	1348	GGCGGATGAGTACAACTGCGGCGCGCGGAGCTTCGCGCGGAGCGTGGATC	1407
CC	of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used	Db	1243	GGGAGGATGAAGAGACGCTGGGAGAGGATGCTTGGAAATTCAAACGCGAGAGTGGATC	1302
CC	in methods of the invention. Note: The sequence data for this patent is	Qy	1408	---AACGAGATGGCGGCTTCGCAACGCTCGCGCTTAAAGTTCAAGTTCAAGGTTCCAGCG	1464
CC	not represented in the printed specification but is based on sequence	Db	1303	TCGCGGACGATGGAATAATTCGTGAATCAGCATCAGTACCGATTCTGTGGGCTTTACGCC	1362
CC	information supplied to Derwent by the European Patent Office	Qy	1465	GGCGCGGAGTCTGCTTGGCAAGGACTCGCGCTACCTGAGATGAAGATGGCGCTGGCC	1524
XX	Sequence 1662 BP; 392 A; 398 C; 481 G; 391 T; 0 U; 0 Other;	Db	1363	GGACCTAGGATCTCTTAGGAAAAGATCTAGCGTATCTGCAGATGAAGACGATCGCGCG	1422
Qy	Query Match 7.8%; Score 148.2; DB 6; Length 1662;	Qy	1525	ATCCTCTTCGCTTCTACAGCTTCGCGCTGTGGAGGGGACCCCGGTGCAATACCGCATG	1584
Db	Best Local Similarity 49.2%; Pred. No. 9e-19;	Db	1423	GCGGTTTATCTCAGGATAGACTAACCGTGGCGCGGACACACAGTGGAGCAGAAGATG	1482
Qy	Matches 670; Conservative 0; Mismatches 598; Indels 93; Gaps 7;	Qy	1585	ATGACCATCTCTCCATGGCGACGCGCTCAAGTCCGGT	1625
Db	277 ACCTACATCGTACCCGGTGAATGTCGAGCATGCTCTCAAGACTAACTTCACCAATTAC	Db	1483	TCGTTGACTTTGTTTCATGAAGAAGGACTTTTGTGTCACCGT	1523
Db	244 ACCGTACGTCGATCCCAAGACATCGAACATGCTCAGACAGGTTGCAACATAC	RESULT 13			
Qy	337 CCCAAGGGAATCGTGTACAGATCCTACATGACGCTGCTTCGCTGACGCGATCTTCAAC	ACL21310/c			
Db	304 CTTAAGGTCTTACGTGSCAAGCGTGTTCATGACTTCTTCGCGCAAGGTATCTTCAAC	ID	ACL21310	standard; DNA; 657 BP.	
Qy	397 GCGACGCGAGCTGTGGAGGAAGCAGAGGACGCGGAGTTCGAGTTCGCTTCCAAG	XX	ACL21310;		
Db	364 TCCGACGCTGACATTTGGCTCTTCCAGCGTAAACCGCGCTCTTGAATTCACACACAGG	XX	27-OCT-2003 (revised)		
Qy	457 AACCTGAGG---GATTTACGCGCATCTGTTTCAGAGAGTACTCCCTGAAGCTGTGCGGT	DT	17-OCT-2003 (first entry)		
Db	424 AGTTGAGGCAAGCATGCTGCTGGTGTGACCGGGGANTCAAGCTCCGTTTGTGCA	XX	DNA clone originating in barley containing SNP encoding sequence #11301.		
Qy	514 ATACTGAGCCAGGATCCAGGACGCAAGTGTGACATGAGGACTTTCATGAGG	XX	Barley; single nucleotide polymorphism; SNP; genotype-phenotype analysis;		
Db	484 ATTTCTGAAAACGCTCAGAAACAATTACGAGCGGTTGATCTCCAAGACTTAATACTACGG	XX	gene; ss.		
Qy	574 ATGAGCTGGATCTCATCTGCAAGTTCGGTTCGGGGTCGAGATCGGACGCTGCGCA	KW			
Db	544 CTCACATTCGACAACTTTGCGGTTTAGCATTCGGTAAAGACACTCGAACCTTGGCAGCG				
Qy	634 GATCTCCCGGAGAACAGCTTCGCGAGCGGTTTCGATCCGCGCAACATCATCATCAGCTG				
Db	604 GGACTTCCCGAAGACGGTTTCGCTCGGTTTCGACCGAGCCACCGAAGCTTCTCTCCAG				
Qy	694 CGGTTTCAT-----CGACCGCTGTGGCGCATCAGAGGTTCTTCCAGCTCGGCTCAGAG				
Db	664 CGGTTTATCTACAGAGTTTCTATGAGGCTGAAGAAATGGTTGGACTCGGCTTAGAA				
Qy	748 GGCCTCCTAGCGCAGAGCATCAAGCTGCTGGAGAGTTTCACTACAGCGTGTATCCGCGG				

Mon Mar 8 10:13:26 2004

XX OS Hordeum vulgare; var. (cul.Akashinriki).

XX PN WO2003057877-A1.

XX PD 17-JUL-2003.

XX PF 16-DEC-2002; 2002WO-IB005403.

XX PR 20-DEC-2001; 2001JP-00387059.

XX PR 20-DEC-2001; 2001JP-00387131.

XX PR 20-DEC-2001; 2001JP-00403299.

XX PR 20-DEC-2001; 2001JP-00403300.

XX PR 27-SEP-2002; 2002JP-00327515.

XX PA (UYN1-) UNIV JAPAN OKAYAMA.

XX PI Sato K, Takeda K, Kohara Y;

XX DR WPI; 2003-587127/55.

XX PT Single nucleotide polymorphism sites in barley varieties and DNA

PT sequences containing them for analysis and identification of barley

PT varieties and production of barley transformants with desired

PT characteristics.

XX PS Disclosure; SEQ ID XX; 284pp; Japanese.

XX CC The present invention relates to oligonucleotide clones originating in

CC barley (Hordeum vulgare) which contain single nucleotide polymorphisms

CC (SNP). The oligonucleotides may be used for analysis of SNPs among barley

CC varieties, identification of particular varieties and genotype-phenotype

CC analysis, isolation of specific genes and creation of new varieties by

CC transformation of barley varieties with them and production of new barley

CC varieties with desired properties. The present sequence represents an

CC oligonucleotide clone DNA sequence featured in the specification. The

CC sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published-pct-sequences. (Updated on 27-OCT-2003 to

CC standardise OS field)

XX SQ Sequence 657 BP; 140 A; 227 C; 159 G; 131 T; 0 U; 0 Other;

Query Match 7.6%; Score 144.6; DB 8; Length 657;

Best Local Similarity 59.3%; Pred. No. 3.9e-18;

Matches 265; Conservative 0; Mismatches 179; Indels 3; Gaps 1;

QY 1182 CGACAGCTCGGCAAGTGTCTACCTCCAGCGTGTCCACGAGACGCTCGCCTGTA 1241

DB 651 CGACAGCTGAGGAGATGCACTACGTGACGCTGCCATCAGGAGTCAATGCACTCTA 592

QY 1242 CCCGCGCTCCCTCAGGACCCCAAGGGGATCCTGGAGGACGAGTGTCTCGGACGGAC 1301

DB 591 CCCGCGCTCGGAGTGAATCGCTGACGCGAGGCGCGAGCTTCTTCGGACGGCAC 532

QY 1302 GAAGGTGAGGCGCGCGGATGTGACGTACGTCCCTACTCGATGGGCGGATGGAGTA 1361

DB 531 GCGSGTGGGACGGTGTTCGTGGCGTCAACTCGTATGCAATGGGCGGATGGAGTC 472

QY 1362 CAATGGGGCCCCGACGGCGGAGCTTCCGCGCGAGGGGTGATCAAC---GAGGATGG 1418

DB 471 TGTGTGGGGCGAGGACCGGAGCGGTATCCGCGCGAGCGGTGGACCGCGGGAGGG 412

QY 1419 CGCGTTCGGCAACCGGTCCGCTTCAAGTTTCAACGGCGTTCCAGCGGGCGGAGGATCTG 1478

DB 411 GACGTTCCGCGCGGAGAGCCCGTTCGGGTATCGGTATCGCGGTTTCCGCGGGCCAAATTTG 352

QY 1479 CCTGGCAAGGACTCGGCGTACCTGCAGATGAAGATGGCGCTGGCCATTCCTTCGCGTT 1538

DB 351 CCTCGAAGGAGATGGGTATATCCAGATGAAGTCTATCGTGGCGTGGTGGAGA 292

QY 1539 CTACAGCTTCGGGTCTGGAGGGGACCCCGGTGACGATGACGATGATGACATCCCTCTC 1598

DB 291 GTTTGAATTGGCGTGGACGGCGGTACCGCGCGGACAGTGGCATCACTCAGTTAOC 232

QY 1599 CATGGCGACGGCTCAAGGTCCGCGT 1625

DB 231 AATGGCGGACGGGTCCCGGTGCGGT 205

## RESULT 14

ID ACL21306/c

XX ACL21306 standard; DNA; 621 BP.

XX AC ACL21306;

XX DT 27-OCT-2003 (revised)

XX DT 17-OCT-2003 (first entry)

XX DE DNA clone originating in barley containing SNP encoding sequence #11297.

XX KW Barley; single nucleotide polymorphism; SNP; genotype-phenotype analysis;

XX KW gene; ss.

XX OS Hordeum vulgare; var. (cul.Akashinriki).

XX PN WO2003057877-A1.

XX PD 17-JUL-2003.

XX PF 16-DEC-2002; 2002WO-IB005403.

XX PR 20-DEC-2001; 2001JP-00387059.

XX PR 20-DEC-2001; 2001JP-00387131.

XX PR 20-DEC-2001; 2001JP-00403299.

XX PR 20-DEC-2001; 2001JP-00403300.

XX PR 27-SEP-2002; 2002JP-00327515.

XX PA (UYN1-) UNIV JAPAN OKAYAMA.

XX PI Sato K, Takeda K, Kohara Y;

XX DR WPI; 2003-587127/55.

XX PT Single nucleotide polymorphism sites in barley varieties and DNA

PT sequences containing them for analysis and identification of barley

PT varieties and production of barley transformants with desired

PT characteristics.

XX PS Disclosure; SEQ ID XX; 284pp; Japanese.

XX CC The present invention relates to oligonucleotide clones originating in

CC barley (Hordeum vulgare) which contain single nucleotide polymorphisms

CC (SNP). The oligonucleotides may be used for analysis of SNPs among barley

CC varieties, identification of particular varieties and genotype-phenotype

CC analysis, isolation of specific genes and creation of new varieties by

CC transformation of barley varieties with them and production of new barley

CC varieties with desired properties. The present sequence represents an

CC oligonucleotide clone DNA sequence featured in the specification. The

CC sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published-pct-sequences. (Updated on 27-OCT-2003 to

CC standardise OS field)

XX SQ Sequence 621 BP; 111 A; 231 C; 180 G; 97 T; 0 U; 2 Other;

Query Match 7.2%; Score 137.2; DB 8; Length 621;

Best Local Similarity 65.3%; Pred. No. 9.7e-17;

Matches 218; Conservative 0; Mismatches 113; Indels 3; Gaps 1;

QY 1182 CGACAGCTCGGCAAGTGTCTACCTCCAGCGTGTCCACGACGCTCGCCTGTA 1241

DB 428 CGACAGCTGAGGAGATGCACTACGTGACGCTGCCATCAGGATCATGCACTCTA 369

QY 1242 CCCGCGCTCCCTCAGGACCCCAAGGGGATCTCTGGAGGACGAGTGTCTCGGACGGAC 1301

Db	368	CCGCGCGTGGCGTGAACCTGCTGACGCGCAGGCGCGGACGTTCTTCCGACGCGCAC	309
QY	1302	GAAGGTGAGCGCGCGGATGCTGACGTACGTCCTACTCGATGGGCGGATGGAGTA	1361
Db	308	GGCGGTGGGACCGGTTGGTTCGTGGGTACAACTCGTATGCAATGGGCGGATGGATC	249
QY	1362	CAACTGGGCGCGGACGCGCGGAGCTTCCGCGGAGCGGTGATCAAC---GAGATGG	1418
Db	248	TGTGTGGGCGAGGACGCGCAGCGTACCGCGGAGCGGTGGCTGACCGCGGAGGG	189
QY	1419	CGGTTCCGCAACGCTGCGCGTTCAGTTCAGCGGTTCCAGCGGCGGCGGAGGATCTG	1478
Db	188	GAGCTTCCGCGCGGAGAGCGCGTTCGGTACATGGCGTTTCACGCGGCGCAAGATTG	129
QY	1479	CCTGGGCAAGACTCGGCGTACCTGTCAGATGAAG	1512
Db	128	CTTCGGAAGAGATGGCGTATATCCAGATGAAG	95
RESULT 15			
ID	ABZ14873	standard; DNA; 1614 BP.	
XX	AC	ABZ14873;	
XX	21-JAN-2003	(first entry)	
DT	XX	Arabidopsis thaliana stress regulated gene SEQ ID NO 2678.	
DE	XX	Arabidopsis thaliana; plant; gene; stress; transgenic; ds.	
XX	KW	Arabidopsis thaliana.	
XX	OS	Arabidopsis thaliana.	
XX	PN	WO200216655-A2.	
XX	PD	28-FEB-2002.	
XX	PF	24-AUG-2001; 2001WO-US026685.	
XX	PR	24-AUG-2000; 2000US-0227866P.	
PR	26-JAN-2001; 2001US-0264647P.		
PR	22-JUN-2001; 2001US-0300111P.		
XX	PA	(SRI ) SCRIPPS RES INST.	
PA	(SYGN ) SYNGENTA PARTICIPATIONS AG.		
XX	PI	Harper JF, Krebs J, Wang X, Zhu T;	
XX	DR	WPI; 2002-304127/34.	
XX	PT	Identifying a stress condition to which a plant cell has been exposed and	
PT	producing plants with increased tolerance to these abiotic stresses.		
XX	PS	Claim 144; SEQ ID NO 2678; 577pp + Sequence Listing; English.	
XX	CC	The invention relates to identifying a stress condition to which a plant	
CC	cell has been exposed, comprising: (a) contacting nucleic acid		
CC	representative of expressed polynucleotides in the plant cell with an		
CC	array or probes representative of the plant cell genome; and (b)		
CC	detecting a profile of expressed polynucleotides in the plant cell		
CC	characteristic of a stress response. The method is useful in the		
CC	production of transgenic plants, cells and seeds and in producing plants		
CC	with increased tolerance to abiotic stress. The present sequence is that		
CC	of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used		
CC	in methods of the invention. Note: The sequence data for this patent is		
CC	not represented in the printed specification but is based on sequence		
CC	information supplied to Derwent by the European Patent Office		
XX	XX	Sequence 1614 BP; 384 A; 365 C; 421 G; 444 T; 0 U; 0 Other;	
QY	Query Match	7.1%; Score 135.6; DB 6; Length 1614;	
XX	Best Local Similarity	48.3%; Pred.No. 2.2e-16;	
XX	Matches 660; Conservative	0; Mismatches 604; Indels 102; Gaps	6;

us-10-021-657-1.rng

Mon Mar 8 10:13:26 2004

Qy	1347	GGGCGGATGAGTACAACTGGGGCCCGACGGCGGAGCTTCGGCCGGAGCGGTGGAT	1406
Db	1236	GGGAGGATGAAATCGACTTGGGAGAGGATTTTGGAAATCAACACAGAGATGGAT	1295
Qy	1407	CAAC-----GAGGATGGCGGTTCCGCAACGCGTCGCGTTCAAGTTACGGCGTTCCA	1460
Db	1296	TTCCGAATCGGACGGTGGGAGATTCAATCATGATCCGTTTAAATTCGTTGCGTTCAA	1355
Qy	1461	GGGCGGCGGAGGATCTGCTGGGCAAGGACTCGCGTACCTGCAGATGAAGATGGCGCT	1520
Db	1356	TGCTGGCCCTAGGATCTCTGGGTAGGATCTGGCGTATCTACAAATGAAATCAATTGC	1415
Qy	1521	GGCCATCCTCTTCGCTTCTACAGCTTCGGGTGCTGGAGGGGCACCGGTGCGATACCG	1580
Db	1416	ATCGGGGTTTTGCTCCGGCACCGTTGACGGTGGTGACGGGGCATAAGGTGGACAGAA	1475
Qy	1581	CATGATGACCATCCTCTCCATGGCGACGGCCTCAAGTCCGCGTC	1626
Db	1476	GATGTCGTTAACTTTATTCATGAAGTACGGTCTTTTAGTTACGTC	1521

Search completed: March 5, 2004, 22:48:59  
Job time : 815 secs

pred. No. is the number of results predicted by chance to have a

RESULT 1	REFERENCES
AX453572	AUTHORS
LOCUS	TITLE
DEFINITION	
ACCESSION	
VERSION	
KEYWORDS	
SOURCE	
ORGANIS	

AX453572	AX453572	1906 bp	DNA	linear	PAT 02-SEP-2002
LOCUS	Sequence 1 from Patent WO0226789.				
COMPLETION					

. Zea mays  
Zea mays  
Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.  
1. Albertsen, M. C., Fox, T., Huffman, G. and Trimmell, M.  
Nucleotide sequences mediating male fertility and method of using

same		100.0%; Score 1906; DB 6; Length 1906;	
JOURNAL	PATENT: WO 0226789-A 1 04-APR-2002;	Query Match	100.0%;
FEATURES	PIONEER HI-BRED INTERNATIONAL, INC. (US)	Best Local Similarity	100.0%;
source	Location/Qualifiers	Matches 1906; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
	1..1906		
	/organism="Zea mays"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:4577"		
	join(<1..1638,1642..1770)		
	/note="unnamed protein product"		
	/codon_start=1		
	/protein_id="CAD38131.1"		
	/db_xref="GI:21712816"		
	/db_xref="REMBL:CAD38131"		
	/translation="EFGTREAHLTPATPSPPFLAGPHKXIALLVLSWILVORNSL		
	RKQKPRWPVIGAVEOLRNVHRMDLVGLSRHTVTVDMPTSTVYIADPNVVE		
	HVLKNTFNYPGIVYRSMOVLGDCIENADGELWRKQRTASPEFASKNLRPSAI		
	VEREYSLKSLGSQAGKVDVMDQELMRMTLSICKVGFVGVIGILSPDLPSNF		
	AQAFDAANIITLRFIDPLMRKFRHVSEALLAQSICKVDFEYTVIRRKABIVE		
	VRAQKQKMKHDIILSRTELGEAGDGGGDKSLRDLVNLNFVIAGRDFTATLSW		
	FTMHAMSHDVAEKRLRELCAPEAREAREEGVTLVLCGADADDFAARVAQFALL		
	TVDSGLKLVYLHACTETLRLYPVPOPKGLEDDLPDGTFKVRAGMVTVPYSWG		
	RMEYNWPDPAASFREMINEDGAFRNASPFKTAFOAGPRICLKQDSNAYLQMKWLA		
	ILFRYSFELLFEGHPVQYHMTLLSMAHLKVRVSRVCHGDLMDIIVPLNPRQITLV		
	LQICMHACKGRKRWVSLVAVLKP"		
ORIGIN			
601	GGGTTCCGGGTCGAGATCGGCACGCTGTCCGAGATCTCCCGGAGAACAGCTTCGCGCAG	660	GGGTTCCGGGTCGAGATCGGCACGCTGTCCGAGATCTCCCGGAGAACAGCTTCGCGCAG
661	GCCTTCGATCCCGCAACATCATCATCACTCGCTCGGGTTATCGACCCGCTGTGCGGCATC	720	GCCTTCGATCCCGCAACATCATCATCACTCGCTCGGGTTATCGACCCGCTGTGCGGCATC
661	GCCTTCGATCCCGCAACATCATCATCACTCGCTCGGGTTATCGACCCGCTGTGCGGCATC	720	GCCTTCGATCCCGCAACATCATCATCACTCGCTCGGGTTATCGACCCGCTGTGCGGCATC
721	AAGAGGTTCTTCCACGTCGGGTGAGAGGCGCTCTCTAGCGAGACATCAAGCTCGTGGAC	780	AAGAGGTTCTTCCACGTCGGGTGAGAGGCGCTCTCTAGCGAGACATCAAGCTCGTGGAC
721	AAGAGGTTCTTCCACGTCGGGTGAGAGGCGCTCTCTAGCGAGACATCAAGCTCGTGGAC	780	AAGAGGTTCTTCCACGTCGGGTGAGAGGCGCTCTCTAGCGAGACATCAAGCTCGTGGAC
781	GAGTTTCACTACAGCTGATCCGCGGAGGAGGAGCCGAGATCGTCGAGGTCGGGCGCAGC	840	GAGTTTCACTACAGCTGATCCGCGGAGGAGGAGCCGAGATCGTCGAGGTCGGGCGCAGC
781	GAGTTTCACTACAGCTGATCCGCGGAGGAGGAGCCGAGATCGTCGAGGTCGGGCGCAGC	840	GAGTTTCACTACAGCTGATCCGCGGAGGAGGAGCCGAGATCGTCGAGGTCGGGCGCAGC
841	GGCAAAACAGGAGAGATGAAGCAGACATCTCTGTCAAGTTTCATCGAGCTGGGCGAGGCC	900	GGCAAAACAGGAGAGATGAAGCAGACATCTCTGTCAAGTTTCATCGAGCTGGGCGAGGCC
841	GGCAAAACAGGAGAGATGAAGCAGACATCTCTGTCAAGTTTCATCGAGCTGGGCGAGGCC	900	GGCAAAACAGGAGAGATGAAGCAGACATCTCTGTCAAGTTTCATCGAGCTGGGCGAGGCC
901	GGCGACGAGCGCGCGCTTCGGGAGCAGATAAGAGCCTCCGGGACGTGCTCAACTTC	960	GGCGACGAGCGCGCGCTTCGGGAGCAGATAAGAGCCTCCGGGACGTGCTCAACTTC
901	GGCGACGAGCGCGCGCTTCGGGAGCAGATAAGAGCCTCCGGGACGTGCTCAACTTC	960	GGCGACGAGCGCGCGCTTCGGGAGCAGATAAGAGCCTCCGGGACGTGCTCAACTTC
961	GTGATCGCGCGCGCGGACACGACCGCGACGACGCTGTCTGTTCACGACATGCGCCATG	1020	GTGATCGCGCGCGCGGACACGACCGCGACGACGCTGTCTGTTCACGACATGCGCCATG
961	GTGATCGCGCGCGCGGACACGACCGCGACGACGCTGTCTGTTCACGACATGCGCCATG	1020	GTGATCGCGCGCGCGGACACGACCGCGACGACGCTGTCTGTTCACGACATGCGCCATG
1021	TCCCAACCGCGAGCTGGCGGAGAGCTGCGCGGAGCTGTGCGGTCGAGCGGAGCGC	1080	TCCCAACCGCGAGCTGGCGGAGAGCTGCGCGGAGCTGTGCGGTCGAGCGGAGCGC
1021	TCCCAACCGCGAGCTGGCGGAGAGCTGCGCGGAGCTGTGCGGTCGAGCGGAGCGC	1080	TCCCAACCGCGAGCTGGCGGAGAGCTGCGCGGAGCTGTGCGGTCGAGCGGAGCGC
1081	GGCGCGAGGAGGCGCTCACGCTGTCTGTCTGCGCGCGGCTGTGACCGCGACGACAGCG	1140	GGCGCGAGGAGGCGCTCACGCTGTCTGTCTGCGCGCGGCTGTGACCGCGACGACAGCG
1081	GGCGCGAGGAGGCGCTCACGCTGTCTGTCTGCGCGCGGCTGTGACCGCGACGACAGCG	1140	GGCGCGAGGAGGCGCTCACGCTGTCTGTCTGCGCGCGGCTGTGACCGCGACGACAGCG
1141	TTTCGCGCGCGCGCTGCGCGAGCTTCGCGGCTCTCTACCTAGCAGAGCTTCGCGAGCTG	1200	TTTCGCGCGCGCGCTGCGCGAGCTTCGCGGCTCTCTACCTAGCAGAGCTTCGCGAGCTG
1141	TTTCGCGCGCGCGCTGCGCGAGCTTCGCGGCTCTCTACCTAGCAGAGCTTCGCGAGCTG	1200	TTTCGCGCGCGCGCTGCGCGAGCTTCGCGGCTCTCTACCTAGCAGAGCTTCGCGAGCTG
1201	GTCTACCTTCCAGCTGCTCACCGGAGCTTCGCGGCTCTCTACCTAGCAGAGCTTCGCGAG	1260	GTCTACCTTCCAGCTGCTCACCGGAGCTTCGCGGCTCTCTACCTAGCAGAGCTTCGCGAG
1201	GTCTACCTTCCAGCTGCTCACCGGAGCTTCGCGGCTCTCTACCTAGCAGAGCTTCGCGAG	1260	GTCTACCTTCCAGCTGCTCACCGGAGCTTCGCGGCTCTCTACCTAGCAGAGCTTCGCGAG
1261	CCCAAGGGGATCTCTGAGGAGCAGCTGTCTGCGGAGCGGACGAGGTGAGGCGCGCGG	1320	CCCAAGGGGATCTCTGAGGAGCAGCTGTCTGCGGAGCGGACGAGGTGAGGCGCGCGG
1261	CCCAAGGGGATCTCTGAGGAGCAGCTGTCTGCGGAGCGGACGAGGTGAGGCGCGCGG	1320	CCCAAGGGGATCTCTGAGGAGCAGCTGTCTGCGGAGCGGACGAGGTGAGGCGCGCGG
1321	ATGGTCACTAGCTGCTTCTGAGGAGCGGAGGTGAGGAGGTGAGGCGCGCGG	1380	ATGGTCACTAGCTGCTTCTGAGGAGCGGAGGTGAGGAGGTGAGGCGCGCGG
1321	ATGGTCACTAGCTGCTTCTGAGGAGCGGAGGTGAGGAGGTGAGGCGCGCGG	1380	ATGGTCACTAGCTGCTTCTGAGGAGCGGAGGTGAGGAGGTGAGGCGCGCGG
1381	GGGAGCTTCGCGCGGAGCGGTGGATCAACGAGGATGGCGCTTCGCGCAACGCTGCGCG	1440	GGGAGCTTCGCGCGGAGCGGTGGATCAACGAGGATGGCGCTTCGCGCAACGCTGCGCG
1381	GGGAGCTTCGCGCGGAGCGGTGGATCAACGAGGATGGCGCTTCGCGCAACGCTGCGCG	1440	GGGAGCTTCGCGCGGAGCGGTGGATCAACGAGGATGGCGCTTCGCGCAACGCTGCGCG
1441	TTCAAGTTTCAAGCGCTTCCAGGCGGGGCGAGGATCTGCTGGGAGAGGAGCTCGGCTAC	1500	TTCAAGTTTCAAGCGCTTCCAGGCGGGGCGAGGATCTGCTGGGAGAGGAGCTCGGCTAC
1441	TTCAAGTTTCAAGCGCTTCCAGGCGGGGCGAGGATCTGCTGGGAGAGGAGCTCGGCTAC	1500	TTCAAGTTTCAAGCGCTTCCAGGCGGGGCGAGGATCTGCTGGGAGAGGAGCTCGGCTAC
1501	CTGAGATGAAGATGGCGCTGCGCATCTCTTCCGCTTCTACAGCTTCGCGCTGCTGGAG	1560	CTGAGATGAAGATGGCGCTGCGCATCTCTTCCGCTTCTACAGCTTCGCGCTGCTGGAG
1501	CTGAGATGAAGATGGCGCTGCGCATCTCTTCCGCTTCTACAGCTTCGCGCTGCTGGAG	1560	CTGAGATGAAGATGGCGCTGCGCATCTCTTCCGCTTCTACAGCTTCGCGCTGCTGGAG
1561	GGGCAACCGGTGAGTACCGCATGATGATCCCTCTCTCATGGCGCAGCGGCTCAAGGTC	1620	GGGCAACCGGTGAGTACCGCATGATGATCCCTCTCTCATGGCGCAGCGGCTCAAGGTC
1561	GGGCAACCGGTGAGTACCGCATGATGATCCCTCTCTCATGGCGCAGCGGCTCAAGGTC	1620	GGGCAACCGGTGAGTACCGCATGATGATCCCTCTCTCATGGCGCAGCGGCTCAAGGTC
1621	CGGCTCTCTAGGCGCGCTCTGATGTCTGAGGAGTTGGATATGATATCGTCCGCTTAAT	1680	CGGCTCTCTAGGCGCGCTCTGATGTCTGAGGAGTTGGATATGATATCGTCCGCTTAAT
1621	CGGCTCTCTAGGCGCGCTCTGATGTCTGAGGAGTTGGATATGATATCGTCCGCTTAAT	1680	CGGCTCTCTAGGCGCGCTCTGATGTCTGAGGAGTTGGATATGATATCGTCCGCTTAAT
1681	CCACGACAAATAACGCTGCTGTGTACAAATTTGATGATGATGATGATGATGATGATGATG	1740	CCACGACAAATAACGCTGCTGTGTACAAATTTGATGATGATGATGATGATGATGATGATG
1681	CCACGACAAATAACGCTGCTGTGTACAAATTTGATGATGATGATGATGATGATGATGATG	1740	CCACGACAAATAACGCTGCTGTGTACAAATTTGATGATGATGATGATGATGATGATGATG









Qy	107	GGATCTCTGGTCCAGAGGTGGAGCCTGAGGAAGAGAAAGCCGCGAGATCATGGCCAGTCA	166
Db	95750	GGATCTCTGGTCCAGAGGTGGAGCCTGAGGAAGAGAAAGCCGCGAGATCATGGCCAGTCA	95809
Qy	167	TCGGCCGCAACGGGTGGAGCAGCTGAGGAACTACACCGGATGCACACTGGGTTGTCGGGT	226
Db	95810	TCGGCGCAGTGGAGCAACTGAAGAACTACCA CAGGATGCATGACTGGCTTGTCTCGAGT	95869
Qy	227	ACCTGTCA CGGCACAGGACAGGTGACCGTTCGACATGCGCGTTCACTTCTTCACCTACATCG	286
Db	95870	ACTTGTGAGGACAGGACCGGTGACCGTTCGACATGCGCTTTCACCTCTACACCTACATTG	95929
Qy	287	CTGACCCGGTGAATGTCGAGCATGTCTCCTCAAGACTAACTTACCACAAATTACCCCAAGGAA	346
Db	95930	CCGACCCGGTGAACGTCGAGCATGTCTCCTGAGACCAACTTACCACAAATTACCCCAAGGTA	95989
Qy	347	-----	346
Db	95990	AAGAACCATAGGATCTTCAGTGTACTGTAAAAATGTTCCTTGCCTTGCACAGTACTAACTGACA	96049
Qy	347	-----TCGTGTACAGATCTTACATGAGCGTCTCTCTC	378
Db	96050	CAAAAAATGTCTGAANAATGCGGGTGAAGTGTACAGGTCTTATGATGTCTGTCTC	96109
Qy	379	GGTGAAGCGCATCTTCAACGCGCAGCGCGAGCTGTGGAGGAACAGAGAAAGACGCGCAGT	438
Db	96110	GGTGATGGCATTTCAATGCCAGCGCGAGATGTGGAGGAAGCAAGGAGACGCGCAGC	96169
Qy	439	TTTCGAGTTTCGCTCCAAGAACCTGAGGAAATTCAGGGCCATTTGTGTTTCAGAGAGTACTCC	498
Db	96170	TTTCGAGTTTCGCTCCAAGAACCTTGAAGAGACTTCAGCACTGTGTGTCTTCAGGAGTACTCC	96229
Qy	499	CTGAAGCTGTCGGGTATCTGAGCCAGGCATCCAGGCAGGCAAAAGTTGTGGACATG---	555
Db	96230	CTGAAGCTATCAAGCAATCTGAGCCAGGCATGCAAGGCGGAGAGTTGTAGACATGCA	96289
Qy	556	-----	555
Db	96290	GTAACCAACTGAATTCCTTGCCTTATACCTAAACATTTCTTGAGAAACCAATTTGTTCCAG	96349
Qy	556	-----CAGGAACTTTACATGAGGATGACGCTGGACTC	587
Db	96350	AATTCGTATGCAAGAACTAACCAAAATTCAGGAATTTTCATGAGGATGACACTGGACTC	96409
Qy	588	CATCTCAAGGTTGGGTTTCGGGTCGAGATCGGCACGCTGTGCGCAGATCTCCCGCAGAA	647
Db	96410	GATCTCAAGGTCGGGTTGGGGTTGAGATCGGACGCTGTCACTGATCTCCCGAGAA	96469
Qy	648	CAGCTTCGCGCAGGCTTCGATGCGCCCAACATCATCATGCTGCGGTTTCATCGACCC	707
Db	96470	CAGCTTTCGCCAGGCATTCGAGCTGCCAACATCATGCTCAGCTGCGGTTTCATCGATCC	96529
Qy	708	GCTGTGCGGCATCAAGAGGTTCTTCCAGTTCGGGTTCAGAGGCGCTCTTAGCGCAGAGCAT	767
Db	96530	TCTGTGGGCTCGAAGAGTTCTTTCAGCTCGGATCAGAGGCTCTCTTCGAGCAGAGCAT	96589
Qy	768	CAGACTCGTGGACGAGTTCACTTACAGCTGATTCGCGGAGGAGGCGGAGATCGTCCA	827
Db	96590	GAAAGCTGGTTGATGACTTTCACCTTACAGCTGATTCGCGCCCGCAAGGCTGAGATCTTGCA	96649
Qy	828	GGTCCGGCCAGCGGCAACAGAGAG-----	853
Db	96650	GGCTCGAGCCAGCGGCAGCAAGCAAGAGAGGTGATCTTCTCTCTGTGCTCAAGAAATCAGT	96709
Qy	854	-----	853
Db	96710	AGAACTGAACCTGACATGGTAAATGGTGATGATCAGATCGGAAAGGTTTGTGTTCTTGATA	96769
Qy	854	-----AGATGAAGCAGCAATCCTGTACGGTTTCATTCGAGCTG	891
Db	96770	TCGTTGATTTGTAATGGCGAGCAGATCAAGACGACATACTGTTCGCGGTTTCAATCGAGCTC	96829

Qy	892	GGCGAGCCCGGACGAC-----GGCGCGCTTCGGGGACGATAAGAGCTCCGGAC	945
Db	96830	GGGGAGCCCGGGGCGACAGAGGGGGCGCGAGCTTCGGGGACGACAAAGAGCTCCGGAC	96889
Qy	946	GTGGTCTCAACTTCGTGATCCCGGGGGGACACGACGGGACACACGCTGCTGCTGTTTC	1005
Db	96890	GTGGTCTCAACTTCGTGATCCCGGGGTGACACGACGGGACGACGCTGCTGCTGTTTC	96949
Qy	1006	ACGCATATGGCATGTCACCCGAGACGTGGCGGAAGAGCTGGCGCGGAGCTGTGCGCG	1065
Db	96950	ACGTATATGGCATGACGACCCGCGCTGCGCGCAAGCTCCGCGGAGCTGGCCGCG	97009
Qy	1066	TTGAGGCGGAGCGCGCGCGAGAGGGGCGTCAAGCTCTGCTCTGGCGGGCGCTGAC	1125
Db	97010	TTGAGAGATGAGCGCGCGCGAGAGGGGCTGGCGT-----CGCGACGCC	97057
Qy	1126	GCGACGACAAAGCGTTGCCCGCGCGGTGGCGCAGTTTCGGGGGCTCTCACTACGAC	1185
Db	97058	GCGCGGAGCGCTGCTTCGCGCGCGCGGTGGCGCAGTTTCGGCTCTGCTAGCTACGAC	97117
Qy	1186	AGCCTCGGAAGTGTCTTACCTCCACGCTCGCTCACGAGACGCTCGGCTGTACCCC	1245
Db	97118	GCGTGGGAAGCTGTGTACTGTACCGGTGCTGACGAGACGCTCCGCTCTTACCCG	97177
Qy	1246	GCGTCCCTCAGACCCCCAAGGGGATCCTGGAGGACGACGTGCTGCCGACGGGACGAAG	1305
Db	97178	GCGTCCCGAGACCCCCAAGGGGATCGTGAGGAGCAGCTGCTCCCGACGGCACCAAG	97237
Qy	1306	GTGAGGCGCGGGATGTGTAGCTACGTGCCCTACTCGATGGGCGGATGGAGTACAAAC	1365
Db	97238	GTGCGGCGCGGGATGTGTAGCTACGTGCCCTACTCGATGGGAGGATGGAGTACAAAC	97297
Qy	1366	TGGGCGCCCGACGCGCGGAGCTTCGCGCGGAGCGGTGGATCA---ACGAGATGGCGCG	1422
Db	97298	TGGGCGCCCGACGCGCGGAGCTTCGCGCGGAGCGGTGGCTCAGCGGCGACGGCGCGCG	97357
Qy	1423	TTCCGCAACCGTTCGCGCTTCAAGTTTCAAGCGGTTCAGAGCGGGCCGAGGATGTGCTG	1482
Db	97358	TTCCGCAACCGTTCGCGCTTCAAGTTTCAAGCGGTTCAGAGCGGGCCGCGGATGTGCTC	97417
Qy	1483	GGCAAGGACTCCGGCTACCTGCAGATGAGATGGCGCTGGCCACTCTTCGGTCTCTTAC	1542
Db	97418	GGCAAGGACTCCGGCTTACCTGCAGATGAGATGGCGCTGGCCACTCTTCGGTCTCTTAC	97477
Qy	1543	AGCTTCGGCTCTGGAGGGGACCCGGTGAGTACCGCATGANGACCATCTCTTCCAATG	1602
Db	97478	ACCTTCGACTCTGCAGAGACACCCCGTCAAGTACCGCATGATGACCATCTCTCCAATG	97537
Qy	1603	GCGACGGCTCAAGTCCGCTCTAGGGCGCTGTA	1641
Db	97538	GCTCACGGCTCAAGTCCGCTCTCCACCTCCGCTGTA	97576
RESULT 5			
AK118803			
LOCUS	AK118803	1695 bp	mRNA
DEFINITION	Arabidopsis thaliana At:g69500	mRNA for unknown protein, complete	linear
ACCESSION	AK118803		
VERSION	AK118803.1	GI:26452619	
KEYWORDS	FLI CDNA; CAP trapper.		
SOURCE	Arabidopsis thaliana (thale cress)		
ORGANISM	Arabidopsis thaliana		
REFERENCE	1		
AUTHORS	Seki, M., Iida, K., Satou, M., Sakurai, T., Akiyama, K., Ishida, J., Nakajima, M., Enju, A., Kamiya, A., Narusaka, M., Carninci, P., Kawai, J., Hayashizaki, Y., and Shinozaki, K.		
TITLE	Arabidopsis thaliana full-length cDNA		
JOURNAL	Published Only in Database (2002)		
REFERENCE	2	(bases 1 to 1695)	

## AUTHORS

Seki, M., Iida, K., Satou, M., Sakurai, T., Akiyama, K., Ishida, J., Nakajima, M., Enjo, A., Kamiya, A., Narusaka, M., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.

TITLE  
JOURNAL

Direct Submission  
Submitted (25-Nov-2002) Motoaki Seki, RIKEN Genomic Sciences Center; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: msekigsc.riken.go.jp)  
URL: <http://pfigweb.gsc.riken.go.jp>, Tel: 81-45-503-9625, Fax: 81-45-503-9386

## COMMENT

An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al. (1998) Plant J. 15:707-720; Seki et al. (2002) Science 296:141-145). cDNA cleaved with BamHI and XhoI was ligated to modified Lambda FLC-1-E vector (Carninci et al. (2001) Genomics 77:79-90) digested with BamHI and SalI. This clone is in a modified plasmid vector.

Please visit our web site (<http://pfigweb.gsc.riken.go.jp/>) for further details.

## FEATURES

location/Qualifiers  
source

1..1695  
/organism="Arabidopsis thaliana"  
/mol\_type="mRNA"  
/cultivar="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="1"  
/clone="RAF121-14-E14"  
/note="common name: thale cress"

## gene

1..1695  
/gene="Atig69500/F10D13\_15"

## CDS

110..1546  
/gene="Atig69500/F10D13\_15"  
/codon\_start=1  
/product="unknown protein"  
/protein\_id="BAC43393.1"  
/db\_xref="GI:26452620"

/translation="MHDLVLYNSRTVVVMPFTTYIADPINVYVLKTNFSY  
PKGTYSYMEVLGDGIFNSDGLWRKQRTAFEFASKNLRDFSTVFKFYSKLF  
TILSQAFKEQVDMQELLMTILDSCIKVGEIVGTLAPLPEHFHAFDTANI  
VTLRFIDPLWKVFLNIGSALLGKSKVNDFTSVIRRKAELEAOISPTNNN  
NNNVKHDILSRILESDPDSEKESLADIVLNVPIAGRDRTATILWYIMNM  
NENVAKLISLQLEKESAEATNSLHQYDFNSFNEKVFEGALLNYSGLKMH  
YLHAFITETLRLYPVQDPKGVLEDDMLPNGTAKVAGMVTYVYPYSGRMEYVWGS  
AALFKPFRWLKDFVQWAFKPFATFOAGPRICLGDKSAVLQKMWMAILCRFYKPHL  
VPNHPVKYRMNTLSMAHLKLVTSRRS"

## ORIGIN

Query Match 28.8%; Score 548.8; DB 8; Length 1695;  
Best Local Similarity 61.3%; Pred. No. 2.8e-59;  
Matches 943; Conservative 0; Mismatches 577; Indels 18; Gaps 3;  
8 TCATGGATCTCTTACACCGATGGGCAAGAAACACAGAGTGTCCAAAGACTTGGCCT 67  
103 TCATGGATCTCTTACACCGATGGGCAAGAAACACAGAGTGTCCAAAGACTTGGCCT 162  
163 GTCATCGCGCAACGGTGGAGCAGTGGAGCACTACCAACCGATGCAAGCTGGCTTGT 222  
68 TTGTTGGAGCAGCGATCGAAGTAACTAATTTGATCGAATGATGATGCTCGTT 127  
223 GGGTACCTGTACCGCAGACAGACAGTACCGTCCACATGCCGTTCACTTCCACCTAC 282  
128 GAGTATCTTTTATACTCAAGAACAGTAGTGGTTCCAAATGCCATTCACTTATATACAT 187  
283 ATCGCTACCGCGTGAATGTCCAGCATGCTCCCTCAAGACTAATCTTCAACCAATCCCAAG 342  
188 ATAGCTGATCCATCAAGTTGATATGCTCTGAACAAACTTCTCCNATTACCCCAAG 247  
343 GGAATCGGTACAGATCTTACATGAGAGTGTCTCTCGGTGACGGCATCTTCAACGCCGAC 402  
248 GGAGAGACTTATCTTCTTATGGAAGTATTTGGGAGATGGATTTTCAATTCGTAT 307  
403 GCGAGCTGTGGAGACAGACAGACAGCGGGAGTTTCGAGTTTCGCTCCAGAACTG 462  
308 GGAGAGCTTTGGAGAAACAGAGAAAAACCGCGAGTTTCGAATTTGCTTCCAAAGATCTT 367

QY 463 AGGATTTTACGCCCATTTGTTTTCAGAGTACTTCCCTGAAGCTGTCCGCTATCTACTGAGC 522  
DB 368 AGAGATTTTATAGTACTGTAGTGTTTAAAGAGTATAGTCTTAAACTCTTTCATCTCCTTTCC 427  
QY 523 CAGCGATCCAAAGCAGGCAAGTTGTGGACATCAGGAACTTTTACATGAGGATGACGCTG 582  
DB 428 CAAGCACTCTTCAAGAGACACAGTAGATATCAGGAATTTGTGATGAGAACTGCTTG 487  
QY 583 GATCCANTCTGCAAGGTTGGTTCGGGTTCGAGATCGGACGCTGTGCGCAGATCTCCCC 642  
DB 488 GATTCAATATGTAAGTGGGATTTGGTGTGAGATAGGACATTTGGCTCCAGAGTTACCA 547  
QY 643 GAGAACAGCTTCGCGAGGCGTTCGATGCGCGCAACATCATCATCAGCTGCGGTTCATC 702  
DB 548 GAGAACTCATTGCTAAGGCTTTTATACCGCAAAATATAATCTAACACTTCGTTTCATA 607  
QY 703 GACCCGCTGTGGCGCATCAAGAGTTCTTCCAGCTCGGGTCAGAGGCCCTCTCTAGGCGAG 762  
DB 608 GATCCTCTTTGGAAGATGAAAAGTTCCTTAAACATAGGCTCTGAGGCAATTTGCTTGA 667  
QY 763 AGATCAAGCTCGTGGACGAGTCACTACAGCGTGTATCCGCGGAGGAAAGCGGAGATC 822  
DB 668 AGCATAAAGTAGTCAATGATTTTCACTATTTCAGTCATAAGAAAGAAAGCAGAGTTA 727  
QY 823 GTGAGG-----TCGGGCGCAGCGCAACACAGAGAGATGAGACGACATCCTG 873  
DB 728 TTAGAGGCAAAATATCTCTCAACAAACAAACAAACAAACAAAGGTGAAGCAC 787  
QY 874 TCAGGTTTACGAGCTGGCGGCGGCGGACGAGCGCGCGGCTTCGGGAGCGATAAG 933  
DB 788 GATATATCTCTCGAGTTTATCGAGATCAGTGCAGTCCAGTAGCAAAAGAAACAGAGAA 847  
QY 934 AGCTCGGACGCTGTGTCTCACTCTGATCCGCGGCGGACGACGCGGCGAGCAG 993  
DB 848 AGCTTAAGAGATATGTCCTCAACTTTGTAATAGTGGAAAGAGATACAAACAGCAAACT 907  
QY 994 CTGTCGTGGTTCACGACATGGCATCTCCACCGGACGTGCCGAGAGCTGCGCGC 1053  
DB 908 CTCACTGGGCTATATACATGATGTAATGAATGAATGTTCTGAGAAACTATATCTCA 967  
QY 1054 GAGCTGTGCGCTTCGAGCGGCGGCGCGCGGAGAGGCGCTCAGCTCGTCTGTC 1113  
DB 968 GAGCTACAAGAACTCGAAGAAAGAAAGCGCGGAGAGGACTAATACATCGTTGCATCAATC 1027  
QY 1114 GCGGCGCTACGCGGACGACAGGCTTCGCGCGCGGCTGGCGAGTTCGCGGCGCTC 1173  
DB 1028 GA-----TACAGAGGATTTCAATTCCTTCAACGAGAGGTAAAGAAATTCGAGAGCTA 1081  
QY 1174 CTCACCTACGACAGCTCGGCAAGCTGTCTACTCCACGCTCGCTCACCGAGAGCTC 1233  
DB 1082 TTGAATCAGATTTCTTAGGAAACTTCACTACTTACATGCTGTGATCAGAGAACTT 1141  
QY 1234 CGCTGTATCCCGCGCTTCCTCAGGACCCCAAGGATCTGAGGACGAGCTGCTGCG 1293  
DB 1142 CGTCTCTACCTCGCTTCCTCAGGATCCGAAAGAGTATTAGAAGATGATATGTTACCA 1201  
QY 1294 GACGGAGCAGAGTGGGCGCGGAGTGTGACGCTACGTCGCTACTCTGATGGGCGG 1353  
DB 1202 AATGGGCAAAAGTAAAGCTGGAGGATGGTCAATATGATCTTACTCAATGGGTCGT 1261  
QY 1354 ATGAGATCAAACTGGGCGCGGAGCTTCGCGCGGAGCGGTGGATCAACGAG 1413  
DB 1262 ATGGAATACATTTGGGATCAGATGAGCATTTGTTTAAACCCGAAAGATGGCTTA---AA 1318  
QY 1414 GATGGCGGCTTCGCAACGCTCGCGTTCAAGTTCAAGGCTTCAGGCGGCGGAGG 1473  
DB 1319 GATGGAGCTTTTCAAAATGATCACCATTCAAGTTCAAGCTTTTTCAGGCTGGACCAAG 1378  
QY 1474 ATCTGCTGGCAAGACTCGGCTTACATGCAATGAGATGGCGTGGCTTCTCTTC 1533  
DB 1379 ATATGTTGGAAAGGATTCAGCTTATCTGCAAAATGAAATGCAATGCAATCTTGC 1438  
QY 1534 CGCTTCTACAGCTTCCGGCTGTGGAGGCGCACCCGCTGCAGTACCGCATGATGACATC 1593

```

Db      1439 AGATTCTATAGTTTCATTGTTGGTGCACAAATCATCTCTGCAAGTATCGGATGATGACGATC 1498
QY      1594 CTCTCCATGGCGCAGCGCTCAAGTCCGCGTCTCTAG 1631
Db      1499 TTATCCATGGCTCATGGCTTGAAGTTACTGTATCCAG 1536

RESULT 6
LOCUS   AX453574
DEFINITION Sequence 3 from Patent WO0226789.
ACCESSION AX453574
VERSION   AX453574.1 GI:21712817
SOURCE   Sorghum bicolor (sorghum)
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
Clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE
AUTHORS Albertsen,M.C., Fox,T., Huffman,G. and Trimmell,M.
TITLE Nucleotide sequences mediating male fertility and method of using
same
JOURNAL Patent: WO 0226789-A 3 04-APR-2002;
PIONEER HI-BRED INTERNATIONAL, INC. (US)
FEATURES
source location/Qualifiers
1..494
/organism="Sorghum bicolor"
/mol_type="unassigned DNA"
/db_xref="taxon:4558"

ORIGIN
Query Match 22.7%; Score 433.2; DB 6; Length 494;
Best Local Similarity 93.7%; Pred. No. 9.2e-45;
Matches 447; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY      259 ATGCGGTTCACCTTCCTACCTACCTACCTGCTGACCGGTGAATGTGAGCATGTCCTCAAG 318
Db      13 ATGCGGTTCACCTTCCTACCTACCTACCTGCTGACCGGTGAATGTGAGCATGTCCTCAAG 72

QY      319 ACTAACTTCACCAATTACCCAGGAATCGTGTACAGATCTCTACATGACGCTGTCCTC 378
Db      73 ACTAACTTCACCAATTACCCAGGGAGCTGTACAGATCTCTACATGATGTCTCTC 132

QY      379 GTGACGGCATCTTCAACGCCAGCGGAGCTGTGGAGGAAGCAGGACGCGAGT 438
Db      133 GTGACGGCATATTCAACGCTGACGCGAGCTGTGGAGGAAGCAGGACGCGAGT 192

QY      439 TTCGAGTTCGCTCCCAAGAACCTGAGGGATTTTCAGGCCATTTGTTTCAGAGGATCTCC 498
Db      193 TTCGAGTTCGCTCCCAAGAACCTGAGGGATTTTCAGGCCATTTGTTTCAGAGGATCTCC 252

QY      499 CTGAAGCTGTGCGGTATCTAGCCAGGATCCCAAGGACGCAAAAGTTGTGACATCGAG 558
Db      253 CTGAAGCTGTGCGGTATCTAGCTCAGGATCCCAAGGACGCAAAAGTTGTGACATCGAG 312

QY      559 GAACCTTTACATGAGATGACGCTGAGCTCACTCTGCAAGGTTGGGTTGGGTTGGGATCG 618
Db      313 GAACCTTTACATGAGATGACGCTGAGCTCACTCTGCAAGGTTGGGTTGGGTTGGGATCG 372

QY      619 GGCAGGCTGCGGATCTCCCGAGACAGCTTCGCGAGCGGTTGATCGCCGAC 678
Db      373 GGCAGGCTGCGGATCTCCCGAGACAGCTTCGCGAGCGGTTGATCGCCGAC 432

QY      679 ATCATCATCAGCTGCGGTTCATTCAGCCCGCTGTGCGCATCAAGAGGTTCTTCCAC 735
Db      433 ATCATCGTCACTGCGGTTCATTCAGCCCGCTGTGCGCATCAAGAGGTTCTTCCAC 489

RESULT 7
AKI19612
LOCUS   AKI19612

```

```

DEFINITION Oryza sativa (japonica cultivar-group) cDNA clone.002-117-G01, full
insert sequence.
ACCESSION AKI19612
VERSION   AKI19612.1 GI:37989235
KEYWORDS  FLI cDNA; oligo capping.
SOURCE    Oryza sativa (japonica cultivar-group)
ORGANISM  Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE
AUTHORS The Rice Full-Length cDNA Consortium, National Institute of
Agrobiological Sciences Rice Full-Length cDNA Project Team.,
Kikuchi,S., Saton,K., Nagata,T., Kawagashira,N., Doi,K.,
Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I.,
Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C.,
Ohtsuki,K., Shishiki,T., Foundation of Advancement of International
Science Genome Sequencing & Analysis Group, Otsu,Y., Murakami,K.,
Iida,Y., Sugano,S., Fujimura,T., Suzuki,Y., Tsunoda,Y.,
Kurosaki,T., Kodama,T., Masuda,H., Kobayashi,M., Xie,Q., Lu,M.,
Narikawa,R., Sugiyama,A., Mizuno,K., Yokomizo,S., Niikura,J.,
Ikeda,R., Ishibiki,J., Kawamata,M., Yoshimura,A., Miura,J.,
Kusumegi,T., Oka,M., Ryu,R., Ueda,M., Matsubara,K., RIKEN,
Kawai,J., Carninci,P., Adachi,J., Aizawa,K., Arakawa,T., Fukuda,S.,
Hara,A., Hashidume,W., Hayatsu,N., Imotani,K., Ishii,Y., Itoh,M.,
Kagawa,I., Kondo,S., Konno,H., Miyazaki,A., Osato,N., Ota,Y.,
Saito,R., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
Yoshino,M. and Hayashizaki,Y.
TITLE Collection, mapping, and annotation of over 28,000 cDNA clones from
japonica rice
JOURNAL Science 301 (5631), 376-379 (2003)
MEDLINE 22752273
PUBMED 12869764
REFERENCE
AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Doi,K.,
Fujimura,T., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W.,
Hayashida,K., Hayashizaki,Y., Hayatsu,N., Hiramoto,K., Hirooka,T.,
Hori,F., Hotta,I., Iida,J., Ikeda,R., Imamura,K., Kikuchi,S.,
Kishimoto,N., Kishikawa-Hirozane,T., Kishimoto,N., Kobayashi,M.,
Kodama,T., Kojima,K., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
Koya,S., Kurihara,C., Kurosaki,T., Kusumegi,T., Li,C., Lu,M.,
Masuda,H., Matsubara,K., Matsuyama,T., Miura,J., Miyazaki,A.,
Mizuno,K., Murakami,K., Murata,M., Nagata,T., Nakamura,Y.,
Nakamura,M., Namiki,T., Narikawa,R., Niikura,J., Nishi,K.,
Nomura,K., Numasaki,R., Ohneda,E., Ohno,M., Ohtsuki,K., Oka,M.,
Ooka,H., Osato,N., Ota,Y., Otsu,Y., Ryu,R., Saitoh,H., Sakai,C.,
Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Satoh,K.,
Shibata,K., Shinagawa,A., Shiraki,T., Shishiki,T., Sugabe,Y.,
Sugano,S., Sugiyama,A., Suzuki,K., Suzuki,Y., Tagami,M.,
Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
Tanaka,T., Tomaru,A., Toya,T., Tsunoda,Y., Ueda,M., Waki,K.,
Xie,Q., Yahagi,W., Yamada,H., Yamamoto,M., Yasunishi,A., Yazaki,J.,
Yokomizo,S. and Yoshimura,A.
TITLE Collection, mapping, and annotation of 28K full-length cDNA clones
from japonica rice
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 1714)
AUTHORS Kikuchi,S.
TITLE Direct Submission
JOURNAL Submitted (31-JAN-2003) Shoshi Kikuchi, National Institute of
Agrobiological Sciences, Department of Molecular Genetics, Head of
Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki
305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp,
Tel:81-29-838-7007, Fax:81-29-838-7007)
COMMENT
This clone is one of the 32K full-length cDNA clones from japonica
rice.
URL : http://cdna01.dna.affrc.go.jp/cdna/
NIAS Rice Full-length cDNA Project Team: Kikuchi,S., Satoh,K.,
Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J.,
Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Namiki,T.,
Ohneda,E., Yahagi,W., Suzuki,K., Li,C., Ohtsuki,K., Shishiki,T.,

```

Yamamoto, M. and Nakahama, Y.  
 FAIS Genome Sequencing & Analysis Group: Ootomo, Y., Iida, Y.,  
 Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M.,  
 Kobayashi, M., Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M.,  
 Masuda, H., Miura, J., Mizuno, K., Nakikawa, R., Tsunoda, Y., Ueda, M.,  
 Ryu, R., Sugano, S., Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M.,  
 Xie, Q., Yokomizo, S., Yoshimura, A., Matsubara, K. and Murakami, K.  
 Genome Exploration Research Group in Riken Genomic Sciences Center  
 and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K.,  
 Akimura, T., Arakawa, T., Carrinchi, P., Fukuda, S., Hanagaki, T.,  
 Hara, A., Hagiwara, W., Hayashida, K., Hayatsu, N., Hiramoto, K.,  
 Hiraoka, T., Hori, F., Iida, J., Imamura, K., Inotani, K., Ishii, Y.,  
 Itoh, M., Kogawa, I., Kanagawa, S., Kato, H., Kawai, J.,  
 Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Komori, H., Kouda, M.,  
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
 Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Osato, N.,  
 Ota, Y., Saitoh, H., Sakai, C., Sakai, K.,  
 Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K.,  
 Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M.,  
 Tagami-Takeda, Y., Tagawa, A., Takahashi, F.,  
 Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K.,  
 Yasunishi, A. and Hayashizaki, Y.

## FEATURES

source  
 1. 1714  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="mRNA"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /clone="002-117-G01"

## ORIGIN

Query Match 19.2%; Score 366.6; DB 8; Length 1714;  
 Best Local Similarity 57.1%; Pred. No. 1.4e-36;  
 Matches 821; Conservative 0; Mismatches 529; Indels 87; Gaps 5;  
 141 GAAAGGCCGAGATCATGGCCAGTCATCGGCGCAACGGTGGAGAGCTGAGGAACCTACCA 200  
 113 GAAGAAAGCAGGATTCGCGCCGTCGGCGGACCATGTTCCACAGTGTCTCACTTCGG 172  
 201 CCGGATCAGCACTGGCTGTTCGGGTACCTGTACGGCAGGACAGTACGCTCGACAT 260  
 173 CCGCTTGTGGAGTACACACGAGGCTCTCCGCAAGTACCGCACTTCGCGATGCTCAC 232  
 261 GCGCTTCACTTCTACACCTACATCGTACCGCCGGTGAATGTTCAGATGTCTCTCAAG 320  
 233 CCGGACCTGCAACTACCTTACACCGCGCGGCGCCAAAGTTCAGACATCTCTCAAGAC 292  
 321 TAACCTTCAACCAATTAACCCCAAGGAATCGTGTACAGATCTTACATGGACGCTCTCTCGG 380  
 293 CAACCTTGCCTAATACGGCAGGCGCCGATGACCCACGAGTGTTCGAGGACCTCTCTCGG 352  
 381 TGACGGCATCTTCAACCGCGAGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 440  
 353 CGACGGGATCTTCAACGTGGAGCGCGGATGTGGCGGAGGAGGAGGAGGAGGAGGAGGAGG 412  
 441 CGAGTTCGCTCCCAAGAACCTGAGGATTTTCAGCGCATTTGTGTTCAGAGTACTCTCT 500  
 413 CGAGTTCCTCCACCGTGTCTGCGGGAGCTACGAGCGCGCGTGTTCGCGGACACCGCGCG 472  
 501 GAAGCTGTTCGGGTATCTG-----AGCCAGGATCTCAAGGAGGAGGAGGAGGAGGAGGAGG 554  
 473 GGAGCTCGCGGCATCTTGAGAGCTGTGTCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGG 532  
 555 GCAGAACTTACATGAGGATGAGCTGGATCTCCATCTGCAAGGTTGGTTGGTTGGTTGGTT 614  
 533 GCAGGATCTGCTGATCGCGGCGAGCTGGAATCTTTCTTCAGGTTGGTTTCGGGGTCAA 592  
 615 GATCGGACGCTGTGCCAGATCTCCCGAGAAC----AGCTTCGCGGAGGAGGAGGAGGAGG 671  
 593 CTTGGGCTGTCTCCCGGATCCAGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 652  
 672 CGCCAAACATCATCAGCTCGGTTTCATCGACCGGCTGTGGCGCATCAAGAGGTTCTT 731

Db 653 CCGGAGCGGAGGAGTGTCTTCCGATTCTTTCGACCTGTCTCGAAGGTCAAGAGGTTTCT 712  
 QY 732 CCACGTCGGGTGAGAGGCGCTCTTAGCGGAGAGCATCAAGCTCTGAGGAGGAGTTCACCTA 791  
 Db 713 CAACATCTCTGTCGGAGGCAACCATGAAGCAGTCGATCCGACCATCAACGACTTCGTGTA 772  
 QY 792 CAGCGTGATCCCGCGGAGGAGGCGGAGATCTGAGGTTCCGGGCGGCGGCGGCAACAGGA 851  
 Db 773 CTCCTATCTGACAGGA-----AGATCGGACAGATGAGCAGAGAGCAACAGAT 823  
 QY 852 GAAGTGAAGACGACGACATCTCTGTCAGGTTCTAGAGCTGGGCGGAGGCGGCGGAGCAG 911  
 Db 824 CGCCAGGAAGAGGACATCTCTGAGGTTCTGCTCGAGAGGAG-----AAGGA 874  
 QY 912 CCGCGCTTCGGGAGGAGTAAGAGCTCCGGAAGCTGCTCAATTCGTTGATCGCCGG 971  
 Db 875 TCCGCGCTCTTCGACACCAAGTACATCAGGAGCATCATACTCACTTCGTCATCGCCGG 934  
 QY 972 GCGGACACGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1031  
 Db 935 CCGGACACGAGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 994  
 QY 1032 CBTGGCCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1091  
 Db 995 GCTACAGGACAGATCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1025  
 QY 1092 GCGGTCAGCTGCTGCTCTGCGGCGGCTGACGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1151  
 Db 1026 -----GACGCCACACCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1054  
 QY 1152 CBTGGCGAGTTCGCGGCGCTCTCCTACCTACGACAGCTCGGCAAGCTGCTCTACCTCCA 1211  
 Db 1055 GCTCAGGATTTCTTCTTCTTCTGACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1114  
 QY 1212 CCGCTCGCTCAGGAGACGCTCGCTGTACCGCGCGCTCTCTCAGGAGGAGGAGGAGGAGGAG 1271  
 Db 1115 CCGAGGTTGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1174  
 QY 1272 CTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1331  
 Db 1175 CTTCTCGGATGACACCTTTCGCGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1234  
 QY 1332 CBTGCGCTACTCGATGGCGGAGTGAATGAGTACAACTGCGGCGCGGAGGAGGAGGAGGAGGAG 1391  
 Db 1235 CCAACGTTACCCATCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1294  
 QY 1392 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1451  
 Db 1295 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1354  
 QY 1452 GCGGTTCCAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1511  
 Db 1355 GCGGTTCCAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1414  
 QY 1512 GATGCGGCTGCGCATCTCTTTCGCTTCTACAGCTTCGCGCTCTCGAGGAGGAGGAGGAGGAG 1568  
 Db 1415 GATGCTTTCGCTGCT 1471

## RESULT 8

AK103182  
 LOCUS 1780 bp mRNA linear PLN 24-JUL-2003  
 DEFINITION Oryza sativa (japonica cultivar-group) cDNA clone:J033121J02, full insert sequence.  
 ACCESSION AK103182  
 VERSION AK103182.1 GI:32988391  
 KEYWORDS FLI CDNA; CAP trapper.  
 SOURCE Oryza sativa (japonica cultivar-group)  
 ORGANISM Oryza sativa (japonica cultivar-group)  
 Rukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzaceae; Oryza.  
 REFERENCE 1

## AUTHORS

The Rice Full-length cDNA Consortium, National Institute of Agrobiological Sciences Rice Full-length cDNA Project Team, Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, K., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuka, K., Shishiki, T., Foundation of Advancement of International Science Genome Sequencing & Analysis Group: Otono, Y., Murakami, K., Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y., Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M., Narikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Niikura, J., Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J., Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN, Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kondo, S., Konno, H., Miyazaki, A., Osato, N., Ota, K., Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Yoshino, M., and Hayashizaki, Y.

Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice

Science 301 (5631), 376-379 (2003)

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

2 (bases 1 to 1780)

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K., Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Hotta, I., Iida, J., Iida, Y., Ikeda, R., Inamura, K., Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M., Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M., Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Kurosaki, T., Kusumegi, T., Li, C., Lu, M., Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A., Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M., Namiki, T., Narikawa, R., Niikura, J., Nishi, K., Nomura, K., Numasaki, R., Ohneda, E., Ohno, M., Ohtsuka, K., Oka, M., Ooka, H., Osato, N., Ota, Y., Otono, Y., Ryu, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K., Shinagawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S., Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami, T., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W., Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S., and Yoshimura, A.

Direct Submission

Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of Agrobiological Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression, 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail: skikuchi@nias.affrc.go.jp, Tel: 81-29-838-7007, Fax: 81-29-838-7007)

This clone is one of the 28k full-length cDNA clones from japonica rice.

URL : <http://cdna01.dna.affrc.go.jp/cDNA/>

NIAS Rice Full-length cDNA Project Team: Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, K., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuka, K., Shishiki, T., and Yamamoto, M.

FATS Genome Sequencing & Analysis Group: Otono, Y., Iida, Y., Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M., Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J., Mizuno, K., Narikawa, R., Niikura, J., Oka, M., Ryu, R., Sugano, S., Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S., Yoshimura, A., Matsubara, K., and Murakami, K.

Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J., Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Osato, N.,

## FEATURES

## source

1. .1780  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="mRNA"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/clone="J033121J02"

## ORIGIN

Query Match 19.2%; Score 366.6; DB 8; Length 1780;  
Best Local Similarity 57.1%; Pred. No. 1.4e-36;  
Matches 821; Conservative 0; Mismatches 529; Indels 87; Gaps 5;

QY 141 GAAGGCCCGAGATCATGGCCAGTCATCCGCGCAACGGTGGAGCAGCTGAGGAACCTACCA 200  
DB 115 GAAGAAACGCAGGTATCCGCCCGTGGCCGCCACCATGTTCCACAGCTGCTCAACTTCGG 174  
QY 201 CCGGATGCAGCACTGGCTTGTCCGGTACCTGTCAACGCGCAGCAGCAGTACCGTCCGACAT 260  
DB 175 CCGTTGCTGGAGTACCAACAGGAGCTCTCCCGAAGTACCGCACCTTCGCGATGCTCAC 234  
QY 261 GCCGTTCACTTCTTACACCTTACATGCGTGACCCGGTGAATGTGAGCATGTTCTTAAGAC 320  
DB 235 CCGGACCTGCAACTACGCTCTACACCGTGCAGCGGCCCACTGCGAGCACAATCTCTCAAGAC 294  
QY 321 TAATCTTCACTTATCCCAAGGAATCGTGATACAGATCTTACATGACCTGCTCTCTCGG 380  
DB 295 CAATCTTCGCAACTACGCGAAGGCCCGGATGACCCAGCAGCTGTGAGGAGCACTCTCTCGG 354  
QY 381 TGACGGCATCTTCAACCGCGACGCGAGCTGTGGAGGAAGCAGAGGAAGACGGCGAGTTT 440  
DB 355 CGACGGGATCTTCAACTGACGGGCGCATGTGGCGGACAGCGCAAGGTTCGCGAGCT 414  
QY 441 CGAGTTCCGCTTCAAGAACTGAGGGATTTTACGCCCATTTGTGTTTACAGAGTACTCTCT 500  
DB 415 CGAGTTCTCCACCGTGTGTGCGGGACTACAGCGCGCGTGTTCGCGCAGCACCGCGCG 474  
QY 501 GAAGCTGTCGGTATATCTG-----ACCGAGCATCCAGGCGGCAAGTGTGGACAT 554  
DB 475 GGAGCTCGCGCGCATCTCTGGAGCGTGTTCGCGCGCGAAGGGCGGAGAGGGTGGACAT 534  
QY 555 GCAGGAACCTTTACATGAGGATGACGCTGGAATCTCCATCTGCAAGTGTGGGTTCGGGTCT 614  
DB 535 GCAGGATCTGCTGATGCGGGCGACGCTGGACTCTTTCTTCAGGGTTGGTTTCGGGGTCAA 594  
QY 615 GATCGGACGCTGTGCGCAGATCTCCCGGAGAAC---AGTTCGCGCAGGGTTCGATGC 671  
DB 595 CATTGGCGTGTCTCTCCGGATCCAGCAAGAGGGGTGTGGTGTTCGCCAGGGCGTTCGACGA 654  
QY 672 CGCCAAATCATCATCAGCTGCGGTTCATCGACCCCTGTGGCGCATCAAGAGTTCCT 731  
DB 655 CGCGAGGAGCAGGTGTGTTCGCAATCTTCGACCTGCTCTGGAAGGTCAAGAGTTCCT 714  
QY 732 CCAGCTCGGGTTCAGAGGGCTCTCTAGCGCAGAGCATCAAGCTCGTGGACGAGTTTCACTA 791  
DB 715 CAACATCTCGTCGAGGCAACCATGAAGCAGTCGATCCGACCATCAACGACTTCGTGTA 774  
QY 792 CAGGCTGATTCGCGGAGGAAGGCCGAGATCTGTGAGGTTCGGGCCAGCGGCACACAGGA 851  
DB 775 CTCATCATCGACAGGA-----AGATCGAAACAGATGAGCAGAGAGCAACACGAATT 825  
QY 852 GAAGATGAAGCAGCAGATCTGTTCACGGTTCATGAGCTGGCGGAGCGCCGCGCAGCGG 911  
DB 826 CGCCAAAGAAAGAGACATCTGTCTCGAGTTTCTGCTCGAGAGGGAG-----AAGGA 876  
QY 912 CGGCGGCTTCGGGAGCAGTAAGAGCCCTCCGGAGCGTGTGTCTCAACTTCGTGATCGCGG 971  
DB 877 TCCCGGCTCTCGAACCAAGTACATCAGGAGCATATCACTCACTTCGTCATCGCGG 936

## TITLE

## JOURNAL

## COMMENT









QY 497 CCTGAAGCTGTCGGGTATATAGAGCCAGGCGATCAAGGAGGAAAGTTGTGACATGC 556  
Db 589 CCGCAAGCTCGCGGCGTCTCCAGCCAGCGCGCTCAACCAATCCATGACATTC 648  
QY 557 AGGAACCTTACATGAGGATGAGCTGGTGAATCCATCTGCAAGTTGGGTTCGGGTGAGA 616  
Db 649 AGGTTTCTTATGAGAGCAAGATGGACTCCATCTTCCACCATCGCGTTCGGCAAGACC 708  
QY 617 TCAGCAGCTGTCGCGAGATCTCCCGAGAGACAGCTTCGCGAGCGGTTCCATCGCGCA 676  
Db 709 TCAACAGCTGCGCGGCTCGCGAGGCGCGCTTCGCGCGCGCTTCGACAGCGCA 768  
QY 677 ACATCATCATCAGCTGGGTTCATGACCCGCTGTGCGCGATCAAGAGTTTTCACAG 736  
Db 769 GCGAGTTTACCATGCTCGCTACCTCAACCCGTTCTGGAAGCTTCGAGGCTCCCTCAACG 828  
QY 737 TCGGTTCAGAGCGCTCTAGCGAGAGATCAAGCTCGTGAGAGTTCACTTACATACAGG 796  
Db 829 TCGGCGCGAGCGATGCTCAAGAGAGGATCAAGGTCTGCGAGGTTCTGTGTACAGC 889  
QY 797 TGATCCGCGGAGGAGCGAGATCGTTCGAGGTCCGCGCGAGCGGCAACAGGAGAGA 856  
Db 889 TCATCCGCTGACAGGTCGCGAGCTCTCCAAACCAAGGCAACGACACTGATTCGAG-- 946  
QY 857 TGAAGCAGCATCTGTTCAGGTTTCATCGAGCTGGGAGGCGCGGAGCGAGCGCGG 916  
Db 947 ----GCAGGATATCTGACAGATTTCAT-----CCAGGCAACGACTAGCGATTCTG 993  
QY 917 GCTTCGGGCGAGTAAGAGCTCCGCGAGCTGCTGCTGCTCACTTCGTCATCGCGGCGG 976  
Db 994 GGACGGTTGATTCAAGTACTGAGAGACATATTTGAACATGTCATAGCGCGAAG 1053  
QY 977 ACAGCGCGGACAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1036  
Db 1054 ACACCAAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1113  
QY 1037 CCGAGAGCTGCGCGGAGCTGTCGCGGTTCGAGGCGGAGCGCGCGGAGGAGGCG 1096  
Db 1114 AGGAGAGATCTGCCACA-----1132  
QY 1097 TCAGCTCTGCTCTGCGGCGGCGCTGACGCGGCGGAGGCGCTTCGCGCGCGGTGG 1156  
Db 1133 -----AGCCATGAGGCGCCACAGCGCGGCGGAGCGGCTTCATCG 1173  
QY 1157 CGAGTTTCGGGCGCTCTCACCTACAGAGCTTCGCGAGGCTTACCTCCAGCGCT 1216  
Db 1174 ACAGGTTCTCGCAGAGCTGACCGAGGCACTGAAACAAGATGCACTATCTGACGCTG 1233  
QY 1217 GCGTACCGAGAGCTCGGCTGTACCCCGCGCTTCCTCAGGACCCCAAGGGATCTGG 1276  
Db 1234 CATGACGAGAGCTCAGGCTATPACCTCAGTTCCACTGGATTAACAGCAGTGTCT 1293  
QY 1277 AGGACGAGCTGCTCGCGGAGGAGAGGAGTGGGCGCGGCGGAGTGGTACGCTACGTCG 1336  
Db 1294 CAGACGATGATTGCCCCAAGGATTCACGCTCAGCAGGAGGACATCGTTCTTACATCC 1353  
QY 1337 CTTACTTCATGGGCGGATGAGTACAACTGGGCGCCCGAGCGGCGAGCTTCGCGCGG 1396  
Db 1354 CTTACGCGATGGGCGGATGAGAGCTTGTGGGCGAAGACGCTGAATCTCTCCGCGCTG 1413  
QY 1397 AGCGGTGATCAAGAGGATGGCGCTTCGCGACGCTCGCGTTCAAGTTACGCGGT 1456  
Db 1414 AAGTTGGCTGATGAAACGCGCTCTTTTACGAGGAGACCGCTTCAATTTACAGCTT 1473  
QY 1457 TCAGGCGGCGGAGGATCTGCTGGCAAGGACTCGCGTACCTGCGAGATGAAGATGG 1516  
Db 1474 TCAGGCGGCGGAGGATCTGCTGGGAGGATTTGCGTACAGGAGATGAAGATCT 1533  
QY 1517 CGTGGCATCTCTTCGCTTCTACAGCTTCCGCTGCTGAGGAGGCGACCGG--GTGC 1573  
Db 1534 TCGCGCGCGTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1593

QY 1574 AGTACCGCATGATGACCATCTCTCTCCATGCGGACGCGCTCAA 1616  
Db 1594 GCTACCGGACCATGATTACACTCTCCGTCGATCAGGCTCTCCA 1636  
RESULT 12  
AK099925  
LOCUS  
DEFINITION  
Oryza sativa (japonica cultivar-group) cDNA clone: J013116M22, full insert sequence.  
ACCESSION  
AK099925.1 GI:32985134  
VERSION  
AK099925  
KEYWORDS  
F11\_CDNA; CAP trapper.  
SOURCE  
Oryza sativa (japonica cultivar-group)  
ORGANISM  
Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.  
REFERENCE  
1  
The Rice Full-Length cDNA Consortium, National Institute of Agricultural Sciences Rice Full-length cDNA Project Team, Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yanagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T., Foundation of Advancement of International Science Genome Sequencing & Analysis Group; Ohtsuki, K., Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y., Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M., Narioka, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Niikura, J., Takeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J., Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN, Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, K., Hara, A., Hashidume, W., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kondo, S., Konno, H., Miyazaki, A., Osato, N., Ota, Y., Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Yoshino, M. and Hayashizaki, Y.  
Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice  
Science 301 (5631), 376-379 (2003)  
22752273  
12869764  
2 (bases 1 to 1934)  
ADACHI, J., AIZAWA, K., AKIMURA, T., ARAKAWA, T., CARNINCI, P., DOI, K., FUJIMURA, T., KAWAGASHIRA, N., KAWAMATA, M., KOBAYASHI, M., KISHIMOTO, N., KODAMA, T., KOJIMA, Y., KONDO, S., KONNO, H., KOUDE, M., KURASAKI, T., KUROSAKI, T., KUSUMEGI, T., LI, C., LU, M., MASUDA, H., MATSUBARA, K., MATSUYAMA, T., MIURA, J., MIYAZAKI, A., MIZUNO, K., MURAKAMI, K., MURATA, M., NAGATA, T., NAKAMURA, M., NAMIKI, T., NARIKAWA, R., NIIKURA, J., NISHI, K., NOMURA, K., NUMASAKI, R., OHNEDA, E., OHNO, M., OHTSUKI, K., OKA, M., OKA, H., OSATO, N., OTA, Y., OTOMO, Y., RYU, R., SAITO, H., SATOH, K., SAKAI, K., SAKAZUME, N., SANO, H., SASAKI, D., SATO, K., SATOH, K., SHIBATA, K., SHINAGAWA, A., SHIRAKI, T., SHISHIKI, T., SOGABE, Y., SUGANO, S., SUGIYAMA, A., SUZUKI, K., SUZUKI, Y., TAGAMI, M., TAGAMI, T., TAGAWA, A., TAKAHASHI, F., TAKAKU-AKASHI, S., TANAKA, T., TOMARU, A., TOYA, T., TSUNODA, Y., UEDA, M., WAKI, K., XIE, Q., YANAGI, W., YAMADA, H., YAMAMOTO, M., YASUNISHI, A., YAZAKI, J., YOKOMIZO, S. and YOSHIMURA, A.  
Direct Submission  
Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of Agricultural Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail: skikuchi@nias.affrc.go.jp, Tel: 81-29-838-7007, Fax: 81-29-838-7007)  
This clone is one of the 28K full-length cDNA clones from japonica rice.  
URL : <http://cdna01.dna.affrc.go.jp/cdna/>  
NIAS Rice Full-length cDNA Project Team: Kikuchi, S., Satoh, K.,







Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Osato, N., Oca, Y., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K., Yasunishi, A. and Hayashizaki, Y.

## FEATURES

Location/Qualifiers  
1..1881  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="mRNA"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/clone="001-036-C03"

## ORIGIN

Query Match 15.4%; Score 293.8; DB 8; Length 1881;  
Best Local Similarity 53.8; Pred. No. 1.6e-27;  
Matches 813; Conservative 0; Mismatches 597; Indels 101; Gaps 6;

```

QY 141 GAAAGGCCCGAGATCATGCGCAGTCATCGCGCAACGGTGGAGCAGCTGAGGAACCTACCA 200
DB 114 GAAGAAACGAGGTATCCGCCCGTGGCGGCACCATGTTCCACAGCTGCTCACTTCGG 173
QY 201 CCGATGCAAGATGCTTGTGGGTACCTGTTCACGGCACAGACAGTGCACCTCGACAT 260
DB 174 CCGCTTCTGGAGTACACACAGGAGTCTCCCGCAAGTACCGCACCTTCGCGATGCTCAC 233
QY 261 GCGTTTCACTTCTACACCTACATCGCTGACCGCGTGAATGTGAGCATGTCTCAAGAC 320
DB 234 CCGACCTGCAATCTACCTCTACACCGTGCAGCGCGGCCAACGTGAGCACATCTCTCAAGAC 293
QY 321 TAACTTCAACAAATTACCCCAAGGAATCGTGTGATCAGATCCATGACGCTGTCTCTCG 380
DB 294 CAACTTCGCCAATACGGCAAGGCCCGATGACCCACGAGTCTGGAGGACCTCTCTCG 353
QY 381 TGAGGGATTTCAACGCCCAACGGCGAGCTGTGGAGGAACAGAGGAAGACGGCGAGTTT 440
DB 354 CGACGGGATTTCAAGCTGTGACGGCGGCGCATGTGGCGGACGCGCAAGGTGCGCAGCTT 413
QY 441 CGAGTTTCGCTTCAAGAACTGAGGATTTTCAGGGCATTTGTTTTCAGAGAGTACTCCCT 500
DB 414 CGAGTTTCTCACCCGTTGTCTCGGGAATACAGCAGCGCGCTGTTTCGCGACACCGCCG 473
QY 501 GAAGCTGTGCGGTATATCTGA-----GCCAGGCATCCAGGCGAGGCAAGTTCTGGACAT 554
DB 474 GGAGCTCGCGGCATCTTGAGCGTGTGCGGGCGGCGAGGGCGGAGAGGTGGACAT 533
QY 555 GCAGGAATTTATCATGAGATGACGCTGAGTCTCACTGCTCAAGGTTGGGTTGGGTCGA 614
DB 534 GCAGGATCTGCTGATGCGGCGGACGCTGGACTCTTTCTTCAGGGTTGGTTTCGGGGTCAA 593
QY 615 GATCGGCAGCTGTGCGGCAGATCTCCCGGAGACAGC---TTGCGCGAGCGTTGATGTC 671
DB 594 CTTTGGCTGTCTCTCGGATCCAGCAAGAGGGCTTGGTGTTCGCGGCGGTTCGACGA 653
QY 672 CGCAACATCATCATCAGCTGCGGTTCATCGACCGCTGTGCGCATCAAGAGGTTCTT 731
DB 654 CGCGAGCGAGCAGGTCTGTTCGATTTCTTGACCTGCTGAGAGTCAAGAGGTTCTT 713
QY 732 CCAGTTCGGGTGAGAGCCCTCTAGCGGAGCATCAAGCTGTGGAAGGTTCACTTA 791
DB 714 CAACATCTCTGCGAGGCAACCATGAGCAGTGCATCCGACCATCAACGACTTCTGTGA 773
QY 792 CAGCGTGATCCGCGGAGGAAGCCAGATCTGTCAGAGTTCGCGGCGGCGGCAACAGGA 851
DB 774 CTCCTATCTGACAGA-----AGATCGACAGATGAGCAGAGACACACGAATT 824
QY 852 GAAGATGAAGCAGCATCTCTGACGGTTTCATCGAGCTGGGCGAGGCGGCGGACGAGCG 911
DB 825 CGCAAGAAAGAGGACATATCTGCGAGGTTTCTGCTCGAGAGGGAG-----AAGGA 875
QY 912 CGCGGCTTCGGGAGCATAAGAGCTTCGGGACGTTGGTCTCACTTCGTGATCGCGG 971

```

```

DB 876 TCCCGCTGCTTCGACAAAGTACATCAGGGAGCATCACTCACTTCGTCATCGCCGG 935
QY 972 GCGGGAACGAGCGGACGACGCTGTCTGGTTTCACGCACATGGCCATGTCCACCGGGA 1031
DB 936 CCGGACACGACGCGGCGGACGCTGTCTGGTTTCCTTACGCGCTCTGCAAGAACCGAGCG 995
QY 1032 CGTGGCCGAGAGCTGCGCGGAGCTGTGGGTTTCGAGGGCGAGCGCGCGCGGAGGA 1091
DB 996 CGTACAGGACAAAGATCGGAGGAAGTGGCGGACGACACACCGCGGACCGCGACCTCG 1055
QY 1092 GCGCGTCACGCTCTGCTCTGCGGGCGCTGACGCGGACGACAAAGCGTTCGCGCGCG 1151
DB 1056 CGTCCAGGATTTCTCTTCACTTCTGACAGAGCGCATCAACAGATCGAGTACCTACA 1115
QY 1152 CGTGGCGAGTTTCGCGGGCTCTCTACCTACGACAGCTTCGGCAA-----1196
DB 1116 CGCAGGTTTGAAGGAGACGCTCCGTTGTACCTGCGCTTCCCTCTGCTGAGTACTGTTCT 1175
QY 1197 -----GCTGCTTACCTC 1209
DB 1176 TGGTCTTTTTTTTTTCGAATAAGAGGCGAGGAGCTGCTACTGTTGTTGCTTGTATT 1235
QY 1210 CAGCGCTCGCTCACCGAGACGCTCCGCTGTACCCCGCGTCCCT-----CAG 1257
DB 1236 TTTTCTTCTGAGGTAGAACTGACGCGCATTTGTTAGTGGCTGTGTTGCGCTGTAG 1295
QY 1258 GACCCCAAGGGATCTCGAGGACGACGCTGCTCCGCGGACGAGTGAAGTGAAGGCGCGC 1317
DB 1296 GATGTCAATACTGCTTCTCGGATGGCAGCTTTCGCGGACGCGGCGCGGTGAAGAAAGA 1355
QY 1318 GGGATGTCACGTCAGTGCCTCTACTCGATGGGCGGATGAGTACAACTGGGGCCCCGAC 1377
DB 1356 GACATGTTGAATACCAACCTGACCCATGCGGAGGATGAAGTTCTGTGCGGCGACGAC 1415
QY 1378 GCGGCGAGCTTCCGCGCGGAGCGTGGATCAACAGGAGTGGCGCTTCGCAACGCTCG 1437
DB 1416 GCGGAGGAGTTCAAGCGCGGAGCGTGGCTCGATGACAGTGGTATGTTCTGCCGAGAGC 1475
QY 1438 CGTTCAAGTTCAAGCGGTTCCAGCGGCGGCGGAGTCTGCTGGGCAAGACTCGGCG 1497
DB 1476 CCGTTCAAGTTCAAGCGGTTCCAGCGGCGGCGGCGGCAAGTCTGTTGGGGAAGAGTTCCGCG 1535
QY 1498 TACCTGCGAGTGAAGTGGCGTGGCCATCTCTTCGCTTCTACAGTTCGCGGCTGCTG 1557
DB 1536 TACGCGAGTGAAGATGTTTCGCTGCTCTCTTCTTCTTCTCAGATTGAGATGGG 1595
QY 1558 GAGGCGGACCC 1568
DB 1596 GATGACGAGCG 1606

```

## RESULT 15

AK069029

LOCUS

DEFINITION

Oryza sativa (japonica cultivar-group) cDNA clone.J023003E06, full

insert sequence.

ACCESSION

AK069029

VERSION

AK069029.1

KEYWORDS

FLI CDNA; CAP trapper.

SOURCE

Oryza sativa (japonica cultivar-group)

ORGANISM

Oryza sativa (japonica cultivar-group)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartoideae; Oryzeae; Oryza.

REFERENCE

1

AUTHORS

The Rice Full-Length cDNA Consortium, National Institute of

Agrobiological Sciences Rice Full-Length cDNA Project Team;

Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K.,

Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I.,

Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C.,

Ohtsuki, K., Shishiki, T., Foundation of Advancement of International

Science Genome Sequencing &amp; Analysis Group, Otsu, Y., Murakami, K.,

Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y.,

Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M., Nariawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Niikura, J., Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J., Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN, Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashidume, W., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kondo, S., Konno, H., Miyazaki, A., Osato, N., Ota, Y., Saito, R., Sasaki, D., Sato, K., Shibata, K., Shingawa, A., Shiraki, T., Yoshino, M., and Hayashizaki, Y.

Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice

Science 301 (5631), 376-379 (2003)

2752273

12869764

2 (bases 1 to 2118)

ADACHI, J., AIZAWA, K., AKIMURA, T., ARAKAWA, T., CARNINCI, P., DOI, K., FUJIMURA, T., FUKUDA, S., HANAGAKI, T., HARA, A., HASHIZUME, W., HAYASHIDA, K., HAYASHIZAKI, Y., HAYATSU, N., HIRAMOTO, K., HIRAOAKA, T., HORI, F., HOTTA, I., IIDA, J., IIDA, Y., IKEDA, R., IMAMURA, K., IMOTANI, K., ISHIBIKI, J., ISHII, Y., ISHIKAWA, M., ITOH, M., KAGAWA, I., KAWAGASHIRA, N., KAWAI, J., KAWAMATA, M., KIKUCHI, S., KISHIKAWA-HIROZANE, T., KISHIMOTO, N., KOBAYASHI, M., KODAMA, I., KOJIMA, K., KOJIMA, Y., KONDO, S., KONNO, H., KOUNO, M., KOYA, S., KURIHARA, C., KUROSAKI, T., KUSUMEGI, T., LI, C., LU, M., MASUDA, H., MATSUBARA, K., MATSUYAMA, T., MIURA, J., MIYAZAKI, A., MIZUNO, K., MURAKAMI, K., MURATA, M., NAGATA, T., NAKAMURA, M., NAMIKI, T., NARIKAWA, R., NIKURA, J., NISHI, K., NAKAMURA, M., NUMASAKI, R., OHNEDA, E., OHNO, M., OHTSUKI, K., OKA, M., OKA, H., OSATO, N., OTA, Y., OTOMO, Y., RYU, R., SATOH, H., SAKAI, C., SAKAI, K., SAKAZUME, N., SANO, H., SASAKI, D., SATO, K., SATOH, K., SHIBATA, K., SHINGAWA, A., SHIRAKI, T., SHISHIKI, T., SOGABE, Y., SUGANO, S., SUGIYAMA, A., SUZUKI, K., SUZUKI, Y., TAGAMI, M., TAGAMI-TAKEDA, Y., TAGAWA, A., TAKAHASHI, F., TAKAKU-AKHIRA, S., TANAKA, T., TOMARU, A., TOYA, T., TSUNODA, Y., UEDA, M., WAKI, K., XIE, Q., YAHAGI, W., YANADA, H., YAMAMOTO, M., YASUNISHI, A., YAZAKI, J., YOKOMIZO, S., and YOSHIMURA, A.

Direct Submission

Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of Agricultural Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression, 2-1-2 Kamondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail: kikuchi@nias.affrc.go.jp, Tel: 81-29-838-7007, Fax: 81-29-838-7007)

This clone is one of the 28K full-length cDNA clones from japonica rice.

URL: <http://cdna01.dna.affrc.go.jp/cDNA/>  
 NIAS Rice Full-length cDNA Project Team: Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yanada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T., and Yamamoto, M.

FAIS Genome Sequencing & Analysis Group: Oromo, Y., Iida, Y., Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M., Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J., Mizuno, K., Nariawa, R., Niikura, J., Oka, M., Ryu, R., Sugano, S., Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S., Yoshimura, A., Matsubara, K., and Murakami, K.

Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J., Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Osato, N., Ota, Y., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akhiru, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K., Yasunishi, A., and Hayashizaki, Y.

Location/Qualifiers

1. 2118

/organism="Oryza sativa (japonica cultivar-group)"

/mol\_type="mRNA"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /clone="J023003E06"

# ORIGIN

Query Match 15.2%; Score 290.2; DB 8; Length 2118;  
 Best Local Similarity 55.9%; Pred. No. 4.3e-27;  
 Matches 755; Conservative 0; Mismatches 503; Indels 93; Gaps 7;

QY	276	CACCTACATCGCTGACCGGTGAATGTCGAGCATGTCTCAAGACTAATCTTCAACAATTA	335
Db	377	CACCGTCACTTCGACCGCGGCAACCTGGAGCACGTCTCAAGTCGGGTTCGACAACTA	436
QY	336	CCCCAAGGAATCGTGTACAGATCTTACATGGAGTGTCTCTCGGTGACGGCATCTTCAA	395
Db	437	CCCCAAGGGGCCCTTCTGGCAGCGCGTCTTCGGGACGCTGTCTGGGACGGCATCTTCAA	496
QY	396	CGCGAGCGGAGTGTGGAG	455
Db	497	CTCCGAGCGGAGAGAGTGGGTTCGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	556
QY	456	GAACCTGAGGGATTTCAAGCCATTTGTTTCAGAGAGTACTCCCTGAA---GCTGTGGG	512
Db	557	CACCTCGGAGCGGCAATGTCCAGTGGTCTTCGCGTCCATCCACCACCGGCTGCTGCC	616
QY	513	TATACTGAGCGAGCATCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	572
Db	617	CATCTCGAGCGAGCG	676
QY	573	GATGACCTGGATCTTCATCTGCAAGTGTGGTTCGGGTTCGGGTTCGGGTTCGGGTTC	632
Db	677	CCTACCTTCGACAAACATCTCGCGCTTCGGGTCGGGTCGGGTCGGGTCGGGTCGG	736
QY	633	AGATCTCCCGAGAGAGAGTTCGCGAGCGGTTCGATGCGCGCCACATCATCATCAGCT	692
Db	737	GGGTTCGCGAGAGAGCGCTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGT	796
QY	693	CGCGTTCATCGAGCGCG---CTGTGGCGCATCAAGAGGTTCCTCCACGTCGGGTGAGA	746
Db	797	CCGCTTCATCTTCCCGAGTACCTGTGGCGCTGCAGAGTGGCTGGGCTTCGGCATGGA	856
QY	747	GGCGCTCTAGCGAGAGAGATCAAGCTCTGAGAGAGTTCACCTACAGCTGATTCGCGCG	806
Db	857	GACCACTGCGCGAGAGAGAGTTCGCGAGCGGTTCGCGAGCGGTTCGCGAGCGGTTC	916
QY	807	GAGGAGGCGAGAGTCTCGAGTTCGGGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAG	866
Db	917	GCGAAGCTCGAG---CTCGCGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	973
QY	867	CATCTGTACAGGTTTCATCGAGTGGGCGAGGCGCGGCGAGGCGGCGGCGGCGGCGG	926
Db	974	CCTGCTGTCCGGTTCATCGGAGAGGCTCTTACTCGAGCG-----1014	
QY	927	CGATAGAGCGCTCGGAGAGTGTGTCTCACTTCGTGATGTCGCGCGGCGGAGACAGCGCG	986
Db	1015	-----AGTCGCTCGACAGTGTGCGCTCACTTCATCTCTCGCGGCGGCGGAGACATCTCT	1069
QY	987	GACGAGCTGTCTGTGTTTCACGCACATGCGCATGTCCTCCACCGCGAGCTGCGCGAGAGCT	1046
Db	1070	CGTGGCGCTCTCTGTGTTTCTTCTGCGTCTGTGTCCACCCACCGCGCGTGGAGCGAGGT	1129
QY	1047	GCGCCGCGAGTGTGCGCTTCGAGGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCG	1106
Db	1130	CGTGACAGAGCTCTGCGC-----CGT	1150
QY	1107	GCTTCGCGCGGCGGTGACGCGAGAGAGGCTTTCGCGCGCGCGCGCGCGCGCGCGCG	1166
Db	1151	GCTCGCGCGCTCCCGCGGCGCGCATGACCGCGCAT-----GTGCGTGGC	1195
QY	1167	GGGCTCTCTCACTACGACAGCTCGGCAAGTGTCTTACCTCCAGCGCTGCGTCAACCA	1226
Db	1196	GGGCGCTTCACTTCGAGAGAGCTCGACAGTCTGGTCTTACCTCAAGGCGGCGCTTCGGA	1255

FEATURES

source



GenCore version S.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2004, 15:25:24 ; Search time 45 Seconds

(without alignments)

4122.771 Million cell updates/sec

Title: US-10-021-657-2

Perfect score: 3079

Sequence: 1 EFGTREALTPATPSPPFL.....ICWHACKGRWYSLVLAWLKP 588

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL.25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query %	Match Length	DB ID	Description
1	2816	91.5	543	10 Q94KE5	Q94KE5 zea mays (m
2	2426.5	78.8	544	10 Q8S7V0	Q8S7V0 oryza sativ
3	1894	61.5	524	10 Q9C788	Q9C788 arabidopsis
4	1771	57.5	478	10 Q8GWJ5	Q8GWJ5 arabidopsis
5	1084	35.2	506	10 Q7XTY4	Q7XTY4 oryza sativ
6	1074.5	34.9	515	10 Q8S7S6	Q8S7S6 oryza sativ
7	1040.5	33.8	495	10 Q8S7S4	Q8S7S4 oryza sativ
8	999.5	32.5	560	10 Q8W372	Q8W372 oryza sativ
9	939.5	32.5	560	10 Q7XD91	Q7XD91 oryza sativ
10	936.5	32.4	545	10 Q7XUR3	Q7XUR3 oryza sativ
11	934.5	32.3	553	10 Q8S9L1	Q8S9L1 arabidopsis
12	988	32.1	516	10 Q8S7S7	Q8S7S7 oryza sativ
13	977.5	31.7	511	10 Q64G31	Q64G31 arabidopsis
14	976	31.7	541	10 Q9FVW8	Q9FVW8 triticum ae
15	971.5	31.6	537	10 Q80823	Q80823 arabidopsis
16	970	31.5	554	10 Q9LMM1	Q9LMM1 arabidopsis

17	970	31.5	554	10 Q8GYT6	Q8GYT6 arabidopsis
18	969.5	31.5	523	10 Q9CAD6	Q9CAD6 arabidopsis
19	961.5	31.2	523	10 Q8LAO9	Q8LAO9 arabidopsis
20	913	29.7	559	10 Q9FMV1	Q9FMV1 arabidopsis
21	908.5	29.5	404	10 Q8H091	Q8H091 oryza sativ
22	907.5	29.5	490	10 Q22162	Q22162 arabidopsis
23	904.5	29.4	550	10 Q9LEY2	Q9LEY2 arabidopsis
24	893.5	29.0	488	10 Q8L7B4	Q8L7B4 arabidopsis
25	879.5	28.6	499	10 Q9LXX7	Q9LXX7 arabidopsis
26	877.5	28.5	513	10 Q8XG36	Q8XG36 arabidopsis
27	874.5	28.4	523	10 Q8H810	Q8H810 oryza sativ
28	870.5	28.3	498	10 Q9LNL3	Q9LNL3 arabidopsis
29	867.5	28.2	511	10 Q8LRA0	Q8LRA0 oryza sativ
30	860.5	27.9	512	10 Q8LRA6	Q8LRA6 oryza sativ
31	852.5	27.7	403	10 Q9SNS7	Q9SNS7 arabidopsis
32	847	27.5	508	10 Q8LGH8	Q8LGH8 arabidopsis
33	846	27.5	511	10 Q8LRA4	Q8LRA4 oryza sativ
34	841.5	27.3	508	10 Q8LRA7	Q8LRA7 oryza sativ
35	841	27.3	510	10 Q9FMV7	Q9FMV7 arabidopsis
36	840	27.3	512	10 Q8W2N3	Q8W2N3 vicia sativ
37	839	27.2	510	10 Q8GWY4	Q8GWY4 arabidopsis
38	837	27.2	522	10 Q9FYLO	Q9FYLO arabidopsis
39	832	27.0	511	10 Q8W2N1	Q8W2N1 arabidopsis
40	831.5	27.0	510	10 Q8W2N4	Q8W2N4 arabidopsis
41	823	26.7	506	10 Q9SMP5	Q9SMP5 arabidopsis
42	813.5	26.4	541	10 Q9LTN1	Q9LTN1 arabidopsis
43	802.5	26.1	511	10 Q8W2N2	Q8W2N2 arabidopsis
44	797.5	25.9	496	10 Q9S833	Q9S833 arabidopsis
45	797.5	25.9	524	10 Q94DD8	Q94DD8 oryza sativ

## ALIGNMENTS

RESULT 1

Q94KE5 PRELIMINARY; PRT; 543 AA.  
ID Q94KE5  
AC Q94KE5  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
DE Cytochrome P450-like protein.  
GN MS-SB200.  
OS Zea mays (Maize).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC PACCAD clade; Panicoideae; Andropogoneae; Zea.  
OX NCBI\_TaxID=4577;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Fox T.W., Trimnell M.R., Albertsen M.C.;  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
DR EMBL: AF366297; AAK52956.1;  
DR GO: GO:0004497; F:monooxygenase activity; IEA.  
DR GO: GO:0006118; P:electron transport; IEA.  
DR InterPro: IPR01128; Cytochrome\_P450.  
DR Pfam: PF00067; P450; 1.  
DR PRINTS: PR00385; P450.  
KW Heme; Monooxygenase; Oxidoreductase.  
SQ SEQUENCE 543 AA; 61308 MW; DB5C3B9D3183CF25 CRC64;

Query Match 91.5%; Score 2816; DB 10; Length 543;  
Best Local Similarity 100.0%; Pred. No. 5.3e-225;  
Matches 541; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 6 EAHLTATPSPPFLAGPHKYIALLVLSWILVQWSLRKQKQPSWPVIGATVEQLRN 65  
Db 3 EAHLTATPSPPFLAGPHKYIALLVLSWILVQWSLRKQKQPSWPVIGATVEQLRN 62  
QY 66 YHRMHWLVGYSRHRHTVVDMEFTSYTYIADPVNVEHLKTNFTNPKGIYVRSYMDVL 125  
Db 63 YHRMHWLVGYSRHRHTVVDMEFTSYTYIADPVNVEHLKTNFTNPKGIYVRSYMDVL 122

QY 126 LGDGFNADGELMRKQKKTASFEFASKNLRDFAIIVFREYSLKLSGILSQASKAGKVDM 185  
 Db 123 LGDGFNADGELMRKQKKTASFEFASKNLRDFAIIVFREYSLKLSGILSQASKAGKVDM 182  
 QY 186 QELYMRMTLDSICKVGFVIGTILSPDLSPNSFAQAFDAANIITLRFIDPLWRIKRFPH 245  
 Db 183 QELYMRMTLDSICKVGFVIGTILSPDLSPNSFAQAFDAANIITLRFIDPLWRIKRFPH 242  
 QY 246 VGSALLAQSIKLVDEFTYVIRRRKAEIIVFRASGQKQKMDHILSRFIELGEAGDGG 305  
 Db 243 VGSALLAQSIKLVDEFTYVIRRRKAEIIVFRASGQKQKMDHILSRFIELGEAGDGG 302  
 QY 306 GFGDDKSLRDVNLNFIAGDRTTATLTSWFTWAMSHPDVAEKLRELCAFEARAREEG 365  
 Db 303 GFGDDKSLRDVNLNFIAGDRTTATLTSWFTWAMSHPDVAEKLRELCAFEARAREEG 362  
 QY 366 VTLVLCGGADDDKAFARVAQFAGLLTYDSLGLVYLHACVTETLRLYPVQDPKGL 425  
 Db 363 VTLVLCGGADDDKAFARVAQFAGLLTYDSLGLVYLHACVTETLRLYPVQDPKGL 422  
 QY 426 EDDVLPDGTKVRAGGMYTYPYSGMRMEYNWGDAAFRPERWINEGDAFRNASPFKFTA 485  
 Db 423 EDDVLPDGTKVRAGGMYTYPYSGMRMEYNWGDAAFRPERWINEGDAFRNASPFKFTA 482  
 QY 486 FOAGPRICLGKDSAYLQMKMALILFRFYSFRLLEGHPVQYRMWMTILSMAGHLKVRVRA 545  
 Db 483 FOAGPRICLGKDSAYLQMKMALILFRFYSFRLLEGHPVQYRMWMTILSMAGHLKVRVRA 542  
 QY 546 V 546  
 Db 543 V 543

## RESULT 2

Q8S7V0 ID Q8S7V0 PRELIMINARY; PRT; 544 AA.  
 AC Q8S7V0;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Putative plant cytochrome P-450 protein.  
 OS OSJNB0091P11.12.  
 GN Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzoae; Oryza.  
 OC NCBI\_TaxID=39947;  
 RN [1]  
 RF SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Buell C.R., Yuan Q., Ouyang S., Liu J., Moffat K.S., Hill J.N.,  
 RA Gansberger K., Brenner M., Burgess S., Hance M., Shvartsbeyn M.,  
 RA Tsitrin T., Riggs F., Hsiao J., Ziemann V., Blunt S., Pai G.,  
 RA Vanaken S.E., Utterback T.R., Feldblyum T.V., Kalb E., Quackenbush J.,  
 RA Salzberg S.L., White O., Fraser C.M.;  
 RT "Oryza sativa chromosome 3 BAC OSJNB0091P11 genomic sequence.";  
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
 DR EMBL; AC073556; AAL84318.1;  
 DR Gramene; Q8S7V0;  
 DR GO; GO:0004497; F:monooxygenase activity; IEA.  
 DR GO; GO:0006118; P:electron transport; IEA.  
 DR InterPro; IPR001128; Cytochrome\_P450.  
 DR Pfam; PF00067; P450; 1.  
 DR PRINTS; PR00385; P450.  
 KW Heme; Monooxygenase; Oxidoreductase.  
 SQ SEQUENCE 544 AA; 61182 MW; 1A52011B956F9EFD CRC64;

Query Match 78.8%; Score 2426.5; DB 10; Length 544;  
 Best Local Similarity 86.4%; Pred. No. 1.2e-192;  
 Matches 471; Conservative 32; Mismatches 31; Indels 11; Gaps 5;

QY 6 EAHLTATPSPFPPLAGPHKYIALLLVLSWILVQWSLRKQKGRSPWPVIGATVEQLRN 65  
 Db 7 EAHAMPVT--SPFPVAGIHKLAIFLVLVLSWILVHKWSLRNQKGRSPWPVIGATVEQLKN 64  
 QY 66 YHRMDHLVGLSRHRTVTVDMPFTSYTYIADPVNVVHVLKTNFTNPKGIIVRSYMDVL 125  
 Db 65 YHRMDHLVGLSRHRTVTVDMPFTSYTYIADPVNVVHVLKTNFTNPKGIIVRSYMDVL 124  
 QY 126 LGDGFNADGELMRKQKKTASFEFASKNLRDFAIIVFREYSLKLSGILSQASKAGKVDM 185  
 Db 125 LGDGFNADGELMRKQKKTASFEFASKNLRDFAIIVFREYSLKLSGILSQASKAGKVDM 184  
 QY 186 QELYMRMTLDSICKVGFVIGTILSPDLSPNSFAQAFDAANIITLRFIDPLWRIKRFPH 245  
 Db 185 QELYMRMTLDSICKVGFVIGTILSPDLSPNSFAQAFDAANIITLRFIDPLWRIKRFPH 244  
 QY 246 VGSALLAQSIKLVDEFTYVIRRRKAEIIVFRASGQKQKMDHILSRFIELGEAGDD-- 303  
 Db 245 VGSALLAQSIKLVDEFTYVIRRRKAEIIVFRASGQKQKMDHILSRFIELGEAGDD-- 304  
 QY 304 GFGDDKSLRDVNLNFIAGDRTTATLTSWFTWAMSHPDVAEKLRELCAFEARARE 363  
 Db 305 GFGDDKSLRDVNLNFIAGDRTTATLTSWFTWAMSHPDVAEKLRELCAFEARARE 364  
 QY 364 EGVTLVLCGGADDDKAFARVAQFAGLLTYDSLGLVYLHACVTETLRLYPVQDPK 422  
 Db 365 EGVTLVLCGGADDDKAFARVAQFAGLLTYDSLGLVYLHACVTETLRLYPVQDPK 419  
 QY 423 GILEDDVLPDGTKVRAGGMYTYPYSGMRMEYNWGDAAFRPERWINEGDAFRNASPF 481  
 Db 420 GILEDDVLPDGTKVRAGGMYTYPYSGMRMEYNWGDAAFRPERWINEGDAFRNASPF 479  
 QY 482 KFTAFQAGPRICLGKDSAYLQMKMALILFRFYSFRLLEGHPVQYRMWMTILSMAGHLKVR 541  
 Db 480 KFTAFQAGPRICLGKDSAYLQMKMALILFRFYSFRLLEGHPVQYRMWMTILSMAGHLKVR 539  
 QY 542 VSRV 546  
 Db 540 VSTSV 544

RESULT 3  
 Q9C788 ID Q9C788 PRELIMINARY; PRT; 524 AA.  
 AC Q9C788;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DE Cytochrome P450, putative.  
 GN F10013.15.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OC NCBI\_TaxID=3702;  
 RN [1]  
 RF SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=21016719; PubMed=11130712;  
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S., S.Y.,  
 RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.,  
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,  
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,  
 RA Dunn P., Egu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,  
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,  
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,  
 RA Kim C.J., Koo H.L., Krenetskaia I., Kurtz D.B., Kwan S., Lam B.,  
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,  
 RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,  
 RA Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,  
 RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,  
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,  
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,









```
||||| 400 DDVLPDGTVPAGSVTISVAGMKVTVGDDCLFEPERWLSADGTFKFBPHDSFRFVA 459
||||| 486 FOAGPRICLGKDSAYLQWK-MALAILFRFYGFRLLEGHPVQYRMVMTILSMHGLKVRV 542
||||| 460 FNAGPRICLGKDLAYLQWNTAGSVLLR-HRLAVAPGHRVQKMSLTILFMKHGLRMEV 516

RESULT 11
Q889L1 ID Q889L1 PRELIMINARY; PRT; 553 AA.
AC Q889L1
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE ATG00360/A_IG005110.21.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Shinn P., Chen H., Cheuk R., Kim C.J., Meyers M.C., Banh J.,
RA Bowser L., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,
RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
RA Seki N., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis cdna clones.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RA Kim C.J., Chen H., Cheuk R., Shinn P., Banh J., Bowser L.,
RA Carninci P., Chang E., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,
RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,
RA Palm C.J., Quach H.L., Sakurai T., Satou M., Seki N., Southwick A.,
RA Tang C.C., Toriumi M., Wu H.C., Yamada K., Yamamura Y., Yu G., Yu S.,
RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Arabidopsis ORF clones.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL: AY074851; AAL75303.1; -.
DR EMBL: AY133339; AM91369.1; -.
DR GO: GO:0004457; F:monooxygenase activity; IEA.
DR GO: GO:0006118; P:electron transport; IEA.
DR InterPro: IPR001128; Cytochrome_P450.
DR Pfam: PF00067; p450; 1.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME P450; 1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 553 AA; 62531 MW; 19A80ED41A7DBC66 CRC64;

Query Match 32.3%; Score 994.5; DB 10; Length 553;
Best Local Similarity 41.0%; Pred. No. 1.1e-73;
Matches 226; Conservative 99; Mismatches 182; Indels 51; Gaps 15;

QY 29 LLLVWLS--WILVQWLSLRKQKPSWFGVIGATVQQLRNHYRHMHWLVGYL----SRHRT 82
Db 8 LLLVAVAAVYVWLTQRTS-RWLKGRVWVPLVLSPLGLIEQDRMHWITENLRACGGTYQT 66
QY 83 VTVDMPFTS-----YTIADPNVNHVLTNTFNFTNPKGIYVRSYMDVLLGDGIFNADGEL 137
Db 67 CICAVPFLAKKGLGVTVTCDFKNIEMKTRFDNPKGPTQAVFHFQGGIFNSDGD 126
QY 138 WRKQRTASPEFASKNLRO-FSAIVFREYSLKSLISQASKAGKGVVDMQELWYRWITLDS 196
Db 127 WLFQRTAALEFTRITLQAMGWNRNGIKLRFCEPILETAQNNEYFVDLQDLILRLTFDN 186
QY 197 ICKVGFGEVIGTSLPDLSPNSPFAQDAANIITLRFIDP--LWRIKRFHFVGSALLAQ 254
```

```
||||| 187 ICGLAFGKGTQRTCAPGLPENGFAFAFDRATEASLQRTFLPBLFRLKWLKGLGVLSLR 246
||||| 255 SIKLVDEFTSVIRRRKAEIVVEVPSAGKQKQKMKH-DILSRFIELGEAGDDGGGDDKSL 313
||||| 247 SLGIDGLDRAVINTRKQELLSQRESGVQ---RHDDLLSRFMKKKQDS-----YSETFL 297
||||| 314 RDVVNLFVIAGRTDTATTLSWFTHMAMSHDPVAEKLRRELCAPEARAREEGVTLVLCGG 373
||||| 298 RHVALNFIAGRTDTSSVALSNFFWLITTHPTVEDKIVREIC-----VLIETR 346
||||| 374 ADADDKAFARVAQFAGLLTYDSLGKLVLYHACVTEITRLYPVAPQPKGILEDDVLFDG 433
||||| 347 TDVS-----SMTAEPLFEFDEVDRLVYLKAALSETLRLYPSPVPEDSKHVVNDILPDG 398
||||| 434 TKVRAGGMVTVVPYSMGMEYNWGPDAASPRPERWIN-EDGAPNAPSPFKETAQAGPRI 492
||||| 399 TFPAGSSVTYSIYAAGRMKSTWGEDCLEFEPKRWISPDGDKFVNHDPQYRFVAFNAGPRI 458
||||| 493 CLGKDSAYLQWK-MALAILFRFYGFRLLEGHPVQYRMVMTILSMHGLKVRVSRVAVCHGDL 551
||||| 459 CLGKDLAYLQWKTIATAAVALLR-HRLTVAPGHEVEQKMSLTILFMKNGLLVNVHKE---DL 513
||||| 552 DMDIVPLNPRQ 562
||||| 514 EVMMKSLVPKE 524

RESULT 12
Q887S7 ID Q887S7 PRELIMINARY; PRT; 516 AA.
AC Q887S7
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Cytochrome P450-like protein.
GN OSJNB0005J14.26.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Buell C.R., Yuan Q., Ouyang S., Liu J., Moffat K.S., Hill J.N.,
RA Gansberger K., Brenner M., Burgess S., Hance M., Shvartsbeyn M.,
RA Tsitrin T., Riggs F., Hsiao J., Zismann V., Blunt S., Pai G.,
RA VanAken S.E., Utterback T.R., Feldblyum T.V., Kalb E., Quackenbush J.,
RA Salzberg S.L., White O., Fraser C.M.;
RL "Oryza sativa chromosome 10 BAC OSJNB0005J14 genomic sequence.";
RN Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA The Rice Chromosome 10 Sequencing Consortium;
RL "In-depth view of structure, activity, and evolution of rice
chromosome 10.";
RN Science 300:1566-1569 (2003).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Buell C.R., Wing R.A., McCombie W.R., Messing J., Yuan Q.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL: AC074232; AM12480.1; -.
DR EMBL: AB017114; AAP54710.1; -.
DR Gramene; Q887S7; -.
DR GO: GO:0004497; F:monooxygenase activity; IEA.
DR GO: GO:0006118; P:electron transport; IEA.
DR InterPro: IPR001128; Cytochrome_P450.
DR Pfam: PF00067; p450; 1.
DR PRINTS: PR00385; P450.
```





us-10-021-657-2.rspt

Mon Mar 8 10:13:28 2004

```

DR PROSITE; PS00086; CYTOCHROME P450; 1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 537 AA; 60916 MW; 95B1A1F0F055E5D4 CRC64;

Query Match      31.6%; Score 971.5; DB 10; Length 537;
Best Local Similarity 40.9%; Pred. No. 8.4e-72;
Matches 217; Conservative 87; Mismatches 182; Indels 45; Gaps 12;

QY 27 IALLVVLVSLVORSLRQKQPRSPFVIGATVEQLRNVRHMDLVGYL-----SRHRT 82
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 7 LMLLSALTAYFLWLTFTSRCLKGRVWPILGSPGLIENCERHMDISDNLRACSGTYQT 66
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 83 VTVDMPETS-----YTIADPVNVEHVLKTNFTNPKGIYVRSYMDVLLGGDIPNADGEL 137
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 67 CICAIPFLAKKQGLVTVTCDPNLEHLKKNFDFNYPKGPTWQAVFHDLLGGQIENSDDGT 126
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 138 WRKQRTASFEFASKNLRFDSA-IVFREYSKLGLSQASKAGKGVDMQELYNRMTLDS 196
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 127 WLFQKTAALFTTRTLRQAMARWNRALKLRFPLILENARLGSEPIDQLLLRLTFDN 186
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 197 ICKYGFGEIGTLSPDLPENSAQAADAAIIITLRFIDP--LWRIKRFPHVGSSEALIAQ 254
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 187 ICGLTFGKDPRTCAPGLPVTFAVAFDRATEASLQRFILPEILMKFKELRLGLEVSLTR 246
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 255 SIKLVDEFTYSVIRRRKAEIIVEVRASGKQKMKHDIILSRFIELGSEAGDDGGFGDDKSIR 314
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 247 SLVQVDNYLSEIITTRKEEMTQHNNKGHH---DDLRSRFIKKES-----YSDETLQ 296
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 315 DVVLNFIAGRDTTATTLVSWFTHMAMSHDPDAEKLRRELCAFEAREEGVTVLVLCGA 374
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 297 RVALNFILAGRDTSSVALSWFFWLITQHPAIEDKILREICTVLVE--TRGDDVALW----- 350
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 375 DADDKAFARVAQFAGLLTYDSLGKLVYLHACVTETRLYPVAVPQDPKGILEDDVLPDGT 434
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 351 -TDEP-----LSCEELDRVLKALSETLRLYPSPVSPDSKRAVKDDVLPDGT 397
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 435 KVRAGGMVTYPVYSMGRMEYNWGPDAASPRPERWINE--GAFRNASPFKFTAPQAGPRI 492
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 398 FVPAGSSITYIYSAGRMKSTWGEDCLEFKPERWISQSDGGRFINHDPFKFVAFNAGPRI 457
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 493 CLGKDSAYLQMK-MALAILFRFYFRILLEGHVPQYRMNTILSMAHGLKYRV 542
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 458 CLGKDLAYLQMKSIASAVLLR-HRLTVVTGHKVEQKMSLTFLMKYGLLVNV 507
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: March 2, 2004, 15:29:11  
Job time : 47 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2004, 15:22:09 ; Search time 18 seconds  
(without alignments)  
1700.958 Million cell updates/sec

Title: US-10-021-657-2

Perfect score: 3079  
Sequence: 1 EFGTEAHLTPATPSPFPL.....ICMHACKGRWVSLVAVLKP 588

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	994.5	32.3	553	1 C862 ARATH	O3066 arabidopsis
2	880.5	28.6	513	1 C861 ARATH	P48422 arabidopsis
3	818.5	26.6	513	1 C942 VICSA	P98188 vicia sativ
4	785.5	25.5	514	1 C941 VICSA	O81117 vicia sativ
5	427	13.9	519	1 CP5V CANAP	P43083 candida api
6	423	13.7	519	1 CP5L DEBHA	Q9V757 debaryomyce
7	411	13.3	519	1 CP5W CANAP	Q12573 candida api
8	406.5	13.2	523	1 CP5C CANMA	P16496 candida mal
9	402	13.1	524	1 CP5F CANTR	P30608 candida tro
10	400.5	13.0	523	1 CP5E CANMA	P24458 candida mal
11	398.5	12.9	531	1 C4E1 DROME	Q9V4T5 drosophila
12	395	12.8	519	1 CP5M CANMA	Q9V758 debaryomyce
13	387	12.6	519	1 CP5C HUMAN	Q12589 candida mal
14	386.5	12.6	524	1 CP5D CANMA	Q12588 candida mal
15	385	12.5	538	1 CP5J CANMA	Q12587 candida tro
16	384.5	12.5	519	1 CP5V CANMA	Q12588 candida mal
17	383.5	12.5	520	1 YRV5 CAEEL	P30607 candida tro
18	381	12.4	522	1 CP5B CANTR	Q27519 caenorhabdi
19	380.5	12.4	518	1 YRVA CAEEL	Q46051 drosophila
20	380	12.3	507	1 C4DE DROME	P98187 homo sapien
21	377.5	12.3	526	1 C4E3 DROME	Q27517 drosophila
22	377	12.2	520	1 CP5H HUMAN	Q09653 caenorhabdi
23	375	12.2	501	1 C4D2 DROME	Q08477 homo sapien
24	372.5	12.1	534	1 YS24 CAEEL	Q27520 caenorhabdi
25	371	12.0	520	1 CP5J HUMAN	P10615 candida tro
26	370.5	12.0	543	1 YRV8 CANTR	Q12585 candida mal
27	370	12.0	543	1 CP5A CANTR	P30610 candida tro
28	369.5	12.0	505	1 CP5T CANMA	Q12586 candida mal
29	368	12.0	517	1 CP5H CANTR	O48921 glycine max
30	367	11.9	521	1 CP5T CANMA	P48323 homo sapien
31	367	11.9	576	1 C972 SOYBN	P11707 oryctolagus
32	366	11.9	520	1 CP5J HUMAN	
33	365	11.9	501	1 CP36 RABIT	

34	362	11.8	503	1 CP39 RAT	P51538 rattus norv
35	360	11.7	526	1 CP55 RAT	P51870 rattus norv
36	359.5	11.7	483	1 C311 DROME	Q9VYQ7 drosophila
37	357	11.6	552	1 C971 PEA	Q43078 pisum sativ
38	356.5	11.6	524	1 CP51 RAT	P33274 rattus norv
39	356	11.6	507	1 CP5G CANTR	P30609 candida tro
40	355.5	11.5	518	1 YRV3 CAEEL	Q27515 caenorhabdi
41	355	11.5	520	1 YRV2 CAEEL	Q27514 caenorhabdi
42	354.5	11.5	510	1 C4DK DROME	Q9W011 drosophila
43	353.5	11.5	510	1 C312 DROME	Q9VYV6 drosophila
44	353	11.5	524	1 CP5B HUMAN	Q9HBI6 homo sapien
45	351.5	11.4	504	1 CP3B MOUSE	Q64459 mus musculu

ALIGNMENTS

RESULT 1

C862 ARATH STANDARD; PRT; 553 AA.

AC O23066;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Cytochrome P450 86A2 (EC 1.14.-.-).  
GN CYP8A2 OR AT4G0360 OR A.IG005110.21 OR F5110.21.  
OS Arabidopsis thaliana (Mouse-ear cross).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC euroids II; Brassicales; Brassicaceae; Arabidopsi.  
OX NCBI\_TaxID=3702;  
[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=cv. Columbia;  
RC Andrews S.;  
RA Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
[2]  
RN SEQUENCE FROM N.A.  
RP STRAIN=cv. Columbia;  
RC MEDLINE=20083488; PubMed=10617198;  
RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,  
Pohl T., Dueterhoeft A., Stiekema W., Entian K.-D., Terryn N.,  
Harris B., Ansoorge W., Brandt P., Grivell L.A., Rieger M., Mueller M.,  
Weichselgartner M., de Simone V., Obermaier B., Mache R., Schmidheini T.,  
Kreiss M., Delsen M., Puigdomenech P., Watson M., Bancroft I.,  
Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,  
Vos P., Hebeisel J., Zimmermann W., Wedler H., Ridley P.,  
Langham S.-A., McCullagh B., Bilham L., Robben J., Vandenbussche P.,  
Van der Schueren J., Grymonprez B., Chuang Y.-J., Aert R., Defoor E.,  
Braeken M., Weltjens I., Voet M., Bastiaens U., Hilbert H., Braun M.,  
Weitzenegger T., Bothe G., Ramsperger U., Hilbert H., Dirks W.,  
Holzer E., Brandt A., Peters S., van Staveren M., Koetter P.,  
Moolman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,  
Berneiser S., Hempel S., Feldpausch M., Iamherth S., Van den Daele H.,  
De Keyser A., Buysshaert C., Gieles J., Villarroel R., De Clercq R.,  
Van Montagu M., Rogers J., Cronin A., Quail M.A., Bray-Allen S.,  
Clark L., Doggett J., Hall S., Kay M., Lennard N., McClay K., Mayes R.,  
Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,  
Borkova D., Bloeker H., Scharfe M., Grimm M., Loehner T.-H.,  
Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,  
Gabel C., Fuchs M., Partmann B., Grandrath K., Dauner D., Herzl A.,  
Neumann S., Agitrou A., Vitale D., Ligouri R., Piravandi E.,  
Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,  
Massenet O., Ogilvie F., Clabaud G., Muendlein A., Felber R.,  
Chedofor T., Cooke R., Berger C., Monfort A., Casacuberta E.,  
Perez-Perez A., Purnelle B., Bent E., Johnson S., facon D., Tesse T.,  
Reijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielke C.,  
Frisman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,  
Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,  
Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,  
Sekhon M., Murray J., Sheet P., Cordes M., Abuthreiden J.,  
Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,

Latrelle P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,  
 Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,  
 Kramer J., Fulton L., Mardis E., Dente M., Pepin K., Hillier L.W.,  
 Nelson J., Spith J., Ryan E., Andrews S., Geisel C., Layman D.,  
 Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,  
 Antonoin B., Zidan M., Strong C., Sun H., Lamar B., Jordan C.,  
 Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,  
 Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,  
 Granat S., Shondy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,  
 Chen E., Marra M.A., Martienssen R., McCombie W.R.;  
 "Sequence and analysis of chromosome 4 of the plant Arabidopsis  
 thaliana";  
 Nature 402:769-777(1999).  
 - SIMILARITY: Belongs to the cytochrome P450 family.  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 DB 29 LLIWVLS--WILVQVSLRQKQPRSPWVIGATVEQLNVRHMDMLVGYL-----SRHRT 82  
 DB 8 LLVVAAYAWLWQFQIS--RWLKGVPWVPLGSLPGLIEQDRDHDWMTENLRACGGTYQT 66  
 QY 83 VTVDMEPTS-----YTVIADPVNVEHVLKTNFTNPKGIVYRSMYDVLGDDGFENADGEL 137  
 DB 67 CIGAVFLAKQGLVTVTCDFPKNLEHMLKTRFDNPKGPTQWQVFDHDFLQGGIFNSDGD 126  
 QY 138 WRKQRTASPEFASKNLRD--PSAIVFREYSIKLSGILSQSKAGKVVDMQELYNWXTLDS 196  
 DB 127 WLFQKTAALFTTTLRQAMGRVWNRGKILRFQCPILETAQNNYEPVDQLDLRLTFDN 186  
 QY 197 ICKVGFGEIGTSPDLSPNSFAQAFDAANIIITLRITDP--LWRIKRFHVGSEALLAQ 254  
 DB 187 ICGLAFGKDTKTCAPGPNPENGAFAPRATASLQRFILPEFLWKLKWLGLGLEVLSLR 246  
 QY 255 SIKLVDEFTYSVIRRRKAEIIVEVRASGKQEKMKH--DILSRFIEGADGGGFGDDKSL 313  
 DB 247 SLGIDGVLDAVINTRKQELLSQRESGVQ---RHDDLRLSRMFKKQDS-----YSETFL 297  
 QY 314 RDVNLVFIAGRTDTATTLSFTHTMAMSHPDVAEKRLRELCAFEARESGVTVLVCGG 373  
 DB 298 RHVALNFTLAGRTSSVALSWFFWLLITHTPTVEDKIVREICS-----VLITRFG 346  
 QY 374 ADADKFAARVAQVAGLTVYDLSGLKVLVHACVTELTLYPAVPODPKGLIEDDVLDPG 433  
 DB 347 TDVS-----SWTAEPLEFDEVDRLVYLKALSETLRLPSPVEDSKHVNDLIDEG 398  
 QY 434 TKVRAGGMVTVYPYSGRMENYNGWPDAAFRPERWIN--EDGAFRNASPFKFTAFQAGPRI 492  
 DB 399 TVPAGSSVTYSIYAAGRMKSTWGEDCLFEKFERWISPDGKFNVDQVRFVAFNAGPRI 458  
 QY 493 CLGKDSAYLQMK--MALAILFRYSFPLELGEHPVQVYMMTILSMAGLKVRSRAVCHGD 551

Query Match 32.3%; Score 994.5; DB 1; Length 553;  
 Best Local Similarity 41.0%; Pred. No. 1.2e-56;  
 Matches 226; Conservative 92; Mismatches 182; Indels 51; Gaps 15;

DB 29 LLIWVLS--WILVQVSLRQKQPRSPWVIGATVEQLNVRHMDMLVGYL-----SRHRT 82  
 DB 8 LLVVAAYAWLWQFQIS--RWLKGVPWVPLGSLPGLIEQDRDHDWMTENLRACGGTYQT 66  
 QY 83 VTVDMEPTS-----YTVIADPVNVEHVLKTNFTNPKGIVYRSMYDVLGDDGFENADGEL 137  
 DB 67 CIGAVFLAKQGLVTVTCDFPKNLEHMLKTRFDNPKGPTQWQVFDHDFLQGGIFNSDGD 126  
 QY 138 WRKQRTASPEFASKNLRD--PSAIVFREYSIKLSGILSQSKAGKVVDMQELYNWXTLDS 196  
 DB 127 WLFQKTAALFTTTLRQAMGRVWNRGKILRFQCPILETAQNNYEPVDQLDLRLTFDN 186  
 QY 197 ICKVGFGEIGTSPDLSPNSFAQAFDAANIIITLRITDP--LWRIKRFHVGSEALLAQ 254  
 DB 187 ICGLAFGKDTKTCAPGPNPENGAFAPRATASLQRFILPEFLWKLKWLGLGLEVLSLR 246  
 QY 255 SIKLVDEFTYSVIRRRKAEIIVEVRASGKQEKMKH--DILSRFIEGADGGGFGDDKSL 313  
 DB 247 SLGIDGVLDAVINTRKQELLSQRESGVQ---RHDDLRLSRMFKKQDS-----YSETFL 297  
 QY 314 RDVNLVFIAGRTDTATTLSFTHTMAMSHPDVAEKRLRELCAFEARESGVTVLVCGG 373  
 DB 298 RHVALNFTLAGRTSSVALSWFFWLLITHTPTVEDKIVREICS-----VLITRFG 346  
 QY 374 ADADKFAARVAQVAGLTVYDLSGLKVLVHACVTELTLYPAVPODPKGLIEDDVLDPG 433  
 DB 347 TDVS-----SWTAEPLEFDEVDRLVYLKALSETLRLPSPVEDSKHVNDLIDEG 398  
 QY 434 TKVRAGGMVTVYPYSGRMENYNGWPDAAFRPERWIN--EDGAFRNASPFKFTAFQAGPRI 492  
 DB 399 TVPAGSSVTYSIYAAGRMKSTWGEDCLFEKFERWISPDGKFNVDQVRFVAFNAGPRI 458  
 QY 493 CLGKDSAYLQMK--MALAILFRYSFPLELGEHPVQVYMMTILSMAGLKVRSRAVCHGD 551

DB 459 CLGKDSAYLQMK--MALAILFRYSFPLELGEHPVQVYMMTILSMAGLKVRSRAVCHGD 551  
 QY 552 DMDIVFLNPRQ 562  
 DB 514 EVMKSLVPKE 524

RESULT 2  
 C861 ARATH  
 ID C861 ARATH STANDARD; PRT; 513 AA.  
 AC P48422; O9FIM3;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Cytochrome P450 86A1 (EC 1.14.--) (CYPLXXXVI) (P450-dependent fatty  
 DE acid omega-hydroxylase).  
 GN CYP86A1 OR CYP86 OR AT5G58860 OR K19W22.6.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OC NCBI TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RA Benveniste I., Durst F.;  
 RT "Cloning, sequencing and expression of CYP86, a new cytochrome P450  
 RT from Arabidopsis thaliana";  
 RL (In) Plant Gene Register PGR95-074.  
 RN [2]  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RX MEDLINE=98162595; PubMed=9500987;  
 RA Benveniste I., Tjiet N., Adas F., Phillips G., Salaun J.P., Durst F.;  
 RT "CYP86A1 from Arabidopsis thaliana encodes a cytochrome P450-dependent  
 RT fatty acid omega-hydroxylase";  
 RL Biochem. Biophys. Res. Commun. 243:688-693(1998).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=99156233; PubMed=10048488;  
 RA Asamizu E., Sato S., Kaneko T., Nakamura Y., Kotani H., Miyajima N.,  
 RA Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. VIII.  
 RT Sequence features of the regions of 1,081,958 bp covered by seventeen  
 RT physically assigned pl and TAC clones";  
 RL DNA Res. 5:379-391(1998).  
 CC -1- FUNCTION: Catalyzes the omega-hydroxylation of various fatty acids  
 CC (FA). Acts on saturated and unsaturated fatty acids with chain  
 CC lengths from C12 to C18 but not on hexadecane.  
 CC -1- SIMILARITY: Belongs to the cytochrome P450 family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X90458; CAA62082.1; -  
 CC EMBL; AB016885; BAB09631.1; -  
 CC PIR; JC5965; JCS965.  
 CC HSP; P14779; IJJP.  
 CC InterPro; IPR001128; Cytochrome\_P450.  
 CC Pfam; PF00067; P450; 1.  
 CC PRINTS; PR00385; P450.  
 CC PROSITE; PS00086; CYTOCHROME\_P450; 1.  
 CC Oxidoreductase; Monooxygenase; Heme; Transmembrane; Multigene family.  
 CC TRANSMEM 7 27 POTENTIAL.  
 CC METAL 456 456 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).  
 CC CONFLICT 387 388 DD -> EH (IN REF. 1 AND 2).  
 CC SEQUENCE 513 AA; 58553 MW; 441A0CBCEB2DB5 CRC64;

Query Match 28.6%; Score 880.5; DB 1; Length 513;  
Best Local Similarity 39.6%; Pred. No. 2.3e-49;  
Matches 210; Conservative 84; Mismatches 181; Indels 55; Gaps 13;

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AF092917; AAC33645.1; --  
CC InterPro; IPR001128; Cytochrome\_P450.  
CC Pfam; PF00067; P450; 1.  
CC PRINTS; PR00385; P450.  
DR PROSITE; PS00086; CYTOCHROME\_P450; FALSE NEG.  
DR Oxidoreductase; Monooxygenase; Heme; Transmembrane;  
KW Endoplasmic reticulum; Multigene family.  
FT TRANSMEM 7 24  
FT METAL 455 455  
FT SEQUENCE 513 AA; 58419 MW; BEA213B45837B1A7 CRC64;  
SQ

Query Match 26.6%; Score 818.5; DB 1; Length 513;  
Best Local Similarity 34.3%; Pred. No. 2.2e-45;  
Matches 188; Conservative 95; Mismatches 190; Indels 75; Gaps 14;

QY 30 LLVLSWLVQR---WSL---RKQK-----PRSWPVIGATVBOILNHRM 69  
DB 3 LETLSWLLFSLSLWFLFLATKTSKPKTPSTTTPKSYPIFGSAFLLANFHR 62  
QY 70 HMLVGYLSR--HRTVTVDMPP--TSYTIADPNVHEVLKTNFTNYPKGIIVRSMDVLL 126  
DB 63 IQWTSILOITPSSITFVLRPFQARQVTAQPAVQVHILRTNFTCYGKGITTFVQSINDFL 122  
QY 127 GGIIFNADGELMRKOKTASFEFASKNLDF--SAIVFREYSLKLSILSOASKAGKVDM 185  
DB 123 GGIIFNADGESMKFQKQISHEFNTRSLKRFVTVVDVLSDELVPVLSQANSQTLLDF 182  
QY 186 QELYMNTLDSICKVGFVGEIGTSPDLSPNSFAQAFDAANIITLR---FIDPLMKR 242  
DB 183 QDILQRLTFDNCIAFGYDPEVLLSLPEIPFAKAFDESSQLSIBLNALIPLLMKVKR 242  
QY 243 FFHVGEALLAQSIKLVDEFTYSVIRRRKAEIVVEVRSAGKQEKMDILSRFTELGEAGD 302  
DB 243 FLNIGVERQLKEAVERGLATKIVNKKELKALQSESESV--DLSRFL----- 293  
QY 303 DGGFGDDKSLRDVNLNFVLAGDATTATLSWTHMAMSHPDVAELRRLRCAFEARAR 362  
DB 294 --SGHSDSEFTDMKVISIILAGRDITSAALTFFWLLSKSHVENEILKEI-----TGK 346  
QY 363 EBGVTLVLCGADADDKAFARVAQFAGLLTYDSGLKLVYLHACVTTTLRLYPVPODPK 422  
DB 347 SEIV-----GYDEVKDMVYTHAALCESMRLLPPLPYDTK 380  
QY 423 GLEDDVLPDGTQKVRAGMVTYTPYSMGRMEYNWGPDAASFRPERWINE-----GAFNA 478  
DB 381 VAVHDDVLPDGTLLVKGMRVTHIYAMGRSEKINGPDWAEFRPRLSLRDEVGKWSFVGI 440  
QY 479 SPFFTAFOAGPRICLKDSAYLQMKWALA--ILPRFYSF--RLLRGHPVQVYMMTILSMH 536  
DB 441 DYISYFVFOAGPRVICGKEMAFLOMKRVAGIMGRFRVVPAMVGEIERPEYTAHTTSVMKG 500  
QY 537 GLKVRVSR 544  
DB 501 GFPVKIEK 508

Query Match 26.6%; Score 880.5; DB 1; Length 513;  
Best Local Similarity 39.6%; Pred. No. 2.3e-49;  
Matches 210; Conservative 84; Mismatches 181; Indels 55; Gaps 13;

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [http://www.isb-sib.ch/annotation/](http://www.isb-sib.ch/announcement/)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AF092917; AAC33645.1; --  
CC InterPro; IPR001128; Cytochrome\_P450.  
CC Pfam; PF00067; P450; 1.  
CC PRINTS; PR00385; P450.  
DR PROSITE; PS00086; CYTOCHROME\_P450; FALSE NEG.  
DR Oxidoreductase; Monooxygenase; Heme; Transmembrane;  
KW Endoplasmic reticulum; Multigene family.  
FT TRANSMEM 7 24  
FT METAL 455 455  
FT SEQUENCE 513 AA; 58419 MW; BEA213B45837B1A7 CRC64;  
SQ

Query Match 26.6%; Score 818.5; DB 1; Length 513;  
Best Local Similarity 34.3%; Pred. No. 2.2e-45;  
Matches 188; Conservative 95; Mismatches 190; Indels 75; Gaps 14;

QY 30 LLVLSWLVQR---WSL---RKQK-----PRSWPVIGATVBOILNHRM 69  
DB 3 LETLSWLLFSLSLWFLFLATKTSKPKTPSTTTPKSYPIFGSAFLLANFHR 62  
QY 70 HMLVGYLSR--HRTVTVDMPP--TSYTIADPNVHEVLKTNFTNYPKGIIVRSMDVLL 126  
DB 63 IQWTSILOITPSSITFVLRPFQARQVTAQPAVQVHILRTNFTCYGKGITTFVQSINDFL 122  
QY 127 GGIIFNADGELMRKOKTASFEFASKNLDF--SAIVFREYSLKLSILSOASKAGKVDM 185  
DB 123 GGIIFNADGESMKFQKQISHEFNTRSLKRFVTVVDVLSDELVPVLSQANSQTLLDF 182  
QY 186 QELYMNTLDSICKVGFVGEIGTSPDLSPNSFAQAFDAANIITLR---FIDPLMKR 242  
DB 183 QDILQRLTFDNCIAFGYDPEVLLSLPEIPFAKAFDESSQLSIBLNALIPLLMKVKR 242  
QY 243 FFHVGEALLAQSIKLVDEFTYSVIRRRKAEIVVEVRSAGKQEKMDILSRFTELGEAGD 302  
DB 243 FLNIGVERQLKEAVERGLATKIVNKKELKALQSESESV--DLSRFL----- 293  
QY 303 DGGFGDDKSLRDVNLNFVLAGDATTATLSWTHMAMSHPDVAELRRLRCAFEARAR 362  
DB 294 --SGHSDSEFTDMKVISIILAGRDITSAALTFFWLLSKSHVENEILKEI-----TGK 346  
QY 363 EBGVTLVLCGADADDKAFARVAQFAGLLTYDSGLKLVYLHACVTTTLRLYPVPODPK 422  
DB 347 SEIV-----GYDEVKDMVYTHAALCESMRLLPPLPYDTK 380  
QY 423 GLEDDVLPDGTQKVRAGMVTYTPYSMGRMEYNWGPDAASFRPERWINE-----GAFNA 478  
DB 381 VAVHDDVLPDGTLLVKGMRVTHIYAMGRSEKINGPDWAEFRPRLSLRDEVGKWSFVGI 440  
QY 479 SPFFTAFOAGPRICLKDSAYLQMKWALA--ILPRFYSF--RLLRGHPVQVYMMTILSMH 536  
DB 441 DYISYFVFOAGPRVICGKEMAFLOMKRVAGIMGRFRVVPAMVGEIERPEYTAHTTSVMKG 500  
QY 537 GLKVRVSR 544  
DB 501 GFPVKIEK 508

RESULT 4  
C941 VICSA  
ID\_C941 VICSA STANDARD; PRT; 514 AA.  
AC O81117;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Cytochrome P450 94A1 (EC 1.14.-.-) (P450-dependent fatty acid

DE DE omega-hydroxylase).  
 GN CYP94A1 OR VAGH111.  
 OS Vicia sativa (Spring vetch) (Rare).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Vicia.  
 OC NCBI\_TaxID=3908;  
 RX [1]  
 RN SEQUENCE FROM N.A.  
 RP TISSUE=Seedling;  
 RC MEDLINE=98264856; PubMed=9601090;  
 RA Tillet N., Helwig C., Pinot F., Le Bouquin R., Lesot A., Durst F.,  
 RA Salauen J.-P., Benveniste I.;  
 RA "Functional expression in yeast and characterization of a clofibrate-  
 RT inducible plant cytochrome P-450 (CYP94A1) involved in cutin monomers  
 RT synthesis";  
 RL Biochem J. 332:583-589 (1998).  
 CC -1- FUNCTION: catalyzes the omega-hydroxylation of various fatty acids  
 CC (FA) from 10 to 18 carbon atoms. The substrate specificity is  
 CC higher for laurate > palmitate > myristate > linolenate >  
 CC linoleate > oleate > caprate. May play a minor role in cutin  
 CC synthesis and could be involved in plant defense.  
 CC -1- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum (By  
 CC similarity).  
 CC -1- INDUCTION: By clofibrate.  
 CC -1- SIMILARITY: Belongs to the cytochrome P450 family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AF030260; AAD10204.1; --  
 CC PIR; T08014; T08014.  
 CC InterPro; IPR001128; Cytochrome\_P450.  
 CC Pfam; PF00067; P450; 1.  
 CC PRINTS; PR00385; P450.  
 CC PROSITE; PS00086; CYTOCHROME\_P450; FALSE NEG.  
 CC Oxidoreductase; Monooxygenase; Heme; Transmembrane;  
 CC Endoplasmic reticulum; Multigene family.  
 CC TRANSMEM 7 29 POTENTIAL.  
 CC FT METAL 458 458 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).  
 CC SEQUENCE 514 AA; 59110 MW; 3D9361380D6C3B0E CRC64;  
 CC  
 CC Query Match 25.5%; Score 785.5; DB 1; Length 514;  
 CC Best Local Similarity 35.0%; Pred. No. 2.9e-43;  
 CC Matches 165; Conservative 95; Mismatches 154; Indels 57; Gaps 8;  
 CC  
 CC 50 PRSPVIGATVQLRNYHRMDWL---VGYLSRHRVTVDMPFTSYTI-ADPVNVEHLK 106  
 CC 48 PKSYELIGSYLSFRKNLHRIQWLSDIVQISPSATFQDGLGRQIITGNPSTVQHILK 107  
 CC 107 TMTNYPGIVYRMYDVLGDGNIENAGELWRKQKTASFEFASKNLDRF-SALVFREY 165  
 CC 108 NOFSNYQGTTFNTLSPFLGFIINTGNPKKFORQVASHFNTKSRNFVEHIVDTL 167  
 CC 166 SLKLSGLISQASKAGKVDMDQELMYRMTLDSICKVGFVEIGTILSPDLPENSFAQFAA 225  
 CC 168 TNRLLPILTSSTQNNILDFQILQFTFDNICNIAFGYDPEYLTPTNRSKFAEAYEDA 227  
 CC 226 NIITLRTFDPL---WRIKRFHVSGEALLAQSLKVLDEFTSVIERRKAEIIVEYRASK 282  
 CC 228 TEISSKRFELPLPIIWKIKFYFNIGSEKRLKAVTEVRSFAKLVREKRELEE-----K 282  
 CC 283 QEKMKHDIILSRTELGEAGDGGGDDKSLRDVNLNFIAGRDITATTLFWTHAMSH 342  
 CC 283 SSLETEMLSRFL-----SSGHDEDFVADIVISFILAGKDTTSAALTFFWLLWKN 334  
 CC 343 POVAEKLRLRELCAFEAREEGVTIVLCGGADADQKAFARVAQFAGLLTYDSIGKLY 402





CC	CONFLICT	52	52	E -> Q (IN REF. 3).		
CC	CONFLICT	147	147	L -> H (IN REF. 3).		
CC	SEQUENCE	523	AA; 59839	MW; 5E5E7EB432DEE30F	CRC64;	
Query Match						
Best Local Similarity 13.2%; Score 406.5; DB 1; Length 523;						
Matches 145; Conservative 71; Mismatches 187; Indels 113; Gaps 22;						
QY	75	GVLRRRTVTVDN	-PFTSYTIA	-----DPNVNVEHLKNTNPKGIIVRSYMD	123	
DB	72	GQLADYDTITFDKY	PNLSSVMTVAGVLKIVTVDENIKAVLATQFNDFALG	-ARAHFD	130	
QY	124	VLLGDGIFNADGELW	RKQRTAFASKNLRDPSAIVFREYSKLISGLISQASK	-AGK	181	
DB	131	PLLGDDGIFTLDGEG	WKLKSRMLRPPQAREQIAHVKAL	---EPHVQ	---ILAKQIKLNGKG	184
QY	182	VDMOELMWTLDISCK	VGVEIGTL	-----SP-DLP-ENSPAQAFDAANIITL	231	
DB	185	TFDLQELFFRFTVDT	ATEFLFGESVHSYDEKLGIPAPNDIPGRENFAEAFNTSQHYLAT	244		
QY	232	RFIDPL-W-RIKRPHV	GEALLAQSIKLVDEFTYSVIRRKAEIVSVRAGSKQKMK	287		
DB	245	RTYSQIFWLTNPKER	DCN-----AKVHKLAQVFNVTALNATEKEY	-----BEKSK	291	
QY	288	HDILSRFIELGEAGD	GGGF-----GDKSLDVDLVNFVIAGRDTTATLISWETH	337		
DB	292	-----GGVFLYELV	KQTRDPKVLQDQLLNINWAGRDITAGLLSPAMF	334		
QY	338	MAMSHPDVAELKREL	CAFEARARESGVTILVLCGGADADKAPAAVAQFAGILLTYDSL	397		
DB	335	ELARNPKIWNKLR	EVNFGLEARDVDE	-----ISFETL	370	
QY	398	GKLVYLHACVTETLR	LPAVPQDPKGILEDVLP-----DGTK	---VRAGMVTYVPSM	449	
DB	371	KCKEYLKAVLNETLR	MPSPVPIERTATRTDTLPRGGKDGNSPIFVFKGSSVVYSYKT	430		
QY	450	GRMEYNKGPDAASR	PERWINEGAFRNASPFKTAQAGRICLGKDSAYLQMKMALAI	509		
DB	431	HLKQFYGEDAYEPR	PERWFEFS---TRKLGWALPFGNGPRICLGGQFALTEASYVIAR	487		
QY	510	LFREYSRLLEGHPVQ	Y---RMMTILSMAHGLKVRVS	543		
DB	488	LAQM---FEHLESKE	DETYPNPKIHLTWNHNGVVIS	521		
RESULT 9						
ID	CP5F	CANTR	STANDARD;	PRT;	524	AA.
DT	01-APR-1993	(Rel. 25, Created)				
DT	01-APR-1993	(Rel. 25, Last sequence update)				
DT	28-FEB-2003	(Rel. 41, Last annotation update)				
DE	Cytochrome P450 52A6	(EC 1.14.14.-) (CYPLI1A6)	(Alkane-inducible P450-ALK3).			
GN	CYP52A6					
OS	Candida tropicalis (Yeast).					
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;					
OC	Saccharomycetales; mitosporic Saccharomycetales; Candida.					
OX	NCBI_TaxID=5482;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=ATCC 750;					
RX	MEDLINE=93090274; PubMed=1457045;					
RA	Seghezzi W., Meili C., Ruffiner R., Kuenzi R., Sanglard D.,					
RA	Fiechter A.;					
RT	"Identification and characterization of additional members of the					
RL	cytochrome P450 multigene family CYP52 of Candida tropicalis.";					
CC	DNA Cell Biol. 11:767-780(1992).					
CC	FUNCTION: Together with an NADPH cytochrome P450 the enzyme system					
CC	catalyzes the terminal hydroxylation as the first step in the					
CC	assimilation of alkanes and fatty acids. Preferentially					
CC	hydroxylates hexadecane.					
CC	!- CATALYTIC ACTIVITY: Hydroxylation of N-alkanes at the terminal					

CC	position.					
CC	!- INDUCTION: By various alkanes.					
CC	!- SIMILARITY: Belongs to the cytochrome P450 family.					
CC	-----					
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration					
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation					
CC	the European Bioinformatics Institute. There are no restrictions on its					
CC	use by non-profit institutions as long as its content is in no way					
CC	modified and this statement is not removed. Usage by and for commercial					
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> )					
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> .					
CC	-----					
CC	EMBL; Z13010; CAA78354.1; .					
DR	PIR; S22972; S22972.					
DR	HSSP; P14779; LJPZ.					
DR	InterPro; IPR001128; Cytochrome_P450.					
DR	Pfam; PF00067; P450; 2.					
DR	PRINTS; PR00385; P450.					
DR	PROSITE; PS00086; CYTOCHROME_P450; 1.					
KW	Electron transport; Oxidoreductase; Monooxygenase; Heme;					
KW	Transmembrane.					
FT	TRANSMEM 17 34 POTENTIAL.					
FT	METAL 472 472 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).					
SQ	SEQUENCE 524 AA; 59827 MW; 3128E3F5B2135749 CRC64;					
Query Match						
Best Local Similarity 13.1%; Score 402; DB 1; Length 524;						
Matches 137; Conservative 70; Mismatches 181; Indels 120; Gaps 20;						
QY	46	KQKG-----PRSPVIGATVEQLRNYRHMDVLGYSLRHRTVTVDMPFTSYTYIADPVNV	101			
DB	67	KVKGNGRLAREWPI--KTFDDYPN-HTFYMSITGALKIVITVQE-----NI	110			
QY	102	EHVLTNFTNPKGIIVRSYMDVLLGDGIFNADGELWKRKKTASFEFASKNLRDPSAIV	161			
DB	111	KAVLATQTFDSLGRHAHFVP-LLGDGIFTLDGEGWKHSRAMLRPQFARDQIGHVKAL-	168			
QY	162	FREYSKLISGLISQASK--AGKVDMOELMYRMTLDSICKVGFGEIGTL-----SP	211			
DB	169	--EPHIQ---ILAKQIKLNGKKTFDIQELPFRFTVDTATPFLFGESVHSLYDEKLGITP	223			
QY	212	--DLP--ENSPAQAFDAANIIITLRFIDPLWRIKRFHFVHVGSEALLAQSTK--LVDEFTYSV	266			
DB	224	NEIPGRDNPATAFNTSOHYLATRYSQTF---YFLNPKFRCDNAKVHLYAKYFNKA	279			
QY	267	IRRRKAEIYEVNRAG-----KQEKMKHILSRFIELGAGDDGGFGDKSLRDVVL	318			
DB	280	LNFTPEEIEEKS KSGVFLYELVKQTR-----DPKVLQDQLL	316			
QY	319	NFVIAGRDITATLISWFTMHAMSHPDVAEKLREL---CAFEARAREGVTILVLCGA	374			
DB	317	NIMVAGRDITAGLLSFAMFELARHPEIWSKLREIEVNFVGESRVEE-----	365			
QY	375	DADDKAFAARVAQFAGLLTYDSLGLVYLHACVTETLRLYPAVQDPKGILEDVVL----	430			
DB	366	-----ITFESLKRCYLKAILNETLRNYPSPVFNKRTATRTDTTLPRGG	408			
QY	431	--PDGTYK---VRAGMVTYVPSMGRMEYNWGPDAASFEPEWINEGAFRNASPEKFTAF	486			
DB	409	GPNGTDPIPIPKGSTVAYIVKTHRELYBYYGKADDFPRWERFEFS-----TKKLGWAYVVF	465			
QY	487	QAGPRICLGKDSAYLQMKMALAILFRFY	514			
DB	466	NGGPRICLGGQFALTEASYVITRLVQMF	493			
RESULT 10						
ID	CP5E	CANNA	STANDARD;	PRT;	523	AA.
AC	P24458;					
DT	01-MAR-1992	(Rel. 21, Created)				
DT	01-MAR-1992	(Rel. 21, Last sequence update)				
DT	28-FEB-2003	(Rel. 41, Last annotation update)				

Cytochrome P450 52A5 (EC 1.14.14.-) (CYPLI1A5) (Alkane-inducible P450-ALK2-A) (CYP52A3-B).

CYP52A5.

Candida maltosa (Yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; mitosporic Saccharomycetales; Candida.

OX NCBI\_TaxID=5479;

[1]

RP SEQUENCE FROM N.A.

RX MEDLINE=9210967; PubMed=1368716;

RA Okuma M., Hiki T., Tanimoto T., Schunck W.H., Mueller H.G.,

RA Yano K., Takagi M.;

RT "Evidence that more than one gene encodes n-alkane-inducible

RT cytochrome P-450s in Candida maltosa, found by two-step gene

RT disruption.";

RL Agric. Biol. Chem. 55:1757-1764 (1991).

RN [2]

RP CHARACTERIZATION.

RX MEDLINE=96311366; PubMed=8713123;

RA Zimmer T., Okuma M., Ohta A., Takagi M., Schunck W.H.;

RA "The CYP52 multigene family of Candida maltosa encodes functionally

RT diverse n-alkane-inducible cytochromes P450.";

RL Biochem. Biophys. Res. Commun. 224:784-789 (1996).

CC -!- FUNCTION: Together with an NADPH cytochrome P450 the enzyme system

CC catalyzes the terminal hydroxylation as the first step in the

CC assimilation of alkanes and fatty acids.

CC -!- CATALYTIC ACTIVITY: Hydroxylation of N-alkanes at the terminal

CC position.

CC -!- INDUCTION: By alkanes.

CC -!- SIMILARITY: Belongs to the cytochrome P450 family.

CC

This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

EMBL: D12475; BAA02041.1; -

DR EMBL: S77461; AAC60531.1; -

DR PIR: JQ1039; JQ1039.

DR HSP: P44779; LPP2.

DR InterPro: IPR001128; Cytochrome\_P450.

DR Pfam: PF00067; P450; 1.

DR PRINTS: PR00385; P450.

DR PROSITE: PS00086; CYTOCHROME P450; 1.

DR Electron transport; Oxidoreductase; Monooxygenase; Heme;

KW Transmembrane.

FT TRANSMEM 17 34 POTENTIAL.

FT METAL 471 471 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).

SQ SEQUENCE 523 AA; 59868 MW; DB357103097F7FE8 CRC64;

Query Match 13.0%; Score 400.5; DB 1; Length 523;

Best Local Similarity 27.0%; Pred. No. 1.4e-18;

Matches 138; Conservative 71; Mismatches 193; Indels 109; Gaps 18;

66 YHRHMDLWGLVSRHRTVTVDMPFTSYTYIADPNVHVLKTNFTNPKGIVRSYMDVL 125

87 HHSFYMTVAGMLKIVLTV-----DPENIKAVLATQFNDPALG-ARHAHFDPL 132

126 LGDGINADELWKRKORTASFPASKNLDFSNVREYSLKSLGILSQASK--AGKVV 183

133 LGDGIPTLDGEGWKHSRAMLRPQFAECIAHVKAL---EPHVQ---VLAKQIKLNGKETF 186

184 DMQELYMRWTLDSCKVGVGVEIGTLSPDL-----PEN-----SFAQAFDAANIIILRF 233

187 DLQELFRFRTVDTEFLFGESVHSLDDEKLGVPFPNNIPGRENFAKETSQHYLATRT 246

234 IDPLMRKRPFVHVGSEALL---AQSIKLVDEFTYSVIRRRKAEIVFVRASG-----K 282

247 YSQMF---YFLTPKPEFRDCNAKVHKLQAYFNKALDASDEVAEKSXGKGVFLYLK 302

283 QEKMKHDLILSRFTLGEAGDGGGDKSLRDVNLVFNFIAGRDITATLISFWTHAMSH 342

303 QTR-----DPKVLQQLNIMVAGRDITAGLISFAMFELARN 339

343 PDVAEKLRELCAFEAREAREEGVTVLVCGGADADKFAARVAQFAGLTYSGLKLVY 402

340 PKIWNKLEBI-----EVNFGLGGEARVDE-----ISFELTKKEY 375

403 LHACVITETILYPAVPDPKGLLEDDVLP-----DGTK-----VRAGGMVTVPVSMGRMEY 454

376 LKAVINETLRMPSVPVNFRTATRTDTLPRGGKDGSPFVFKGSSVTVTVYKTHLEE 435

455 NWGPDAAFRPERMINEDGAFRNASPEKFTAFQAGPRICLGKDSAYLQMKMALAILFRFY 514

436 YYGKDAYEFPFRWFEP-----TRKLGWAYVFPFNGPRICLGQGFALTEASYVITRLAQM- 491

515 SFRLEGHVQVY--RMVTILSMAGHLKVRVS 543

492 -FEHLESKDQETYPNKCCHLTMHNEGWFIS 521

RESULT 11

C4E1 DROME

ID C4E1 DROME STANDARD; PRT; 531 AA.

AC Q9V4T5;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Probable cytochrome P450 4el (EC 1.14.-.-) (CYP1V1).

GN CYP4E1 OR CG2062.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Berkely;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abril J.F., Agrayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Brokstein P., Brottier P.,

RA Borkova D., Botchan M.R., Bouck J., Butler H., Cadieu E., Center A., Chandra I.,

RA Burtis K.C., Busam D.A., Butler H., Davenport L.B., Davies P.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Dew I., Dietz S.M.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Liukkonen M., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacleb J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195(2000).  
CC -!- FUNCTION: May be involved in the metabolism of insect hormones and  
CC in the breakdown of synthetic insecticides (By similarity).  
CC -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +  
CC oxidized flavoprotein + H(2)O.  
CC -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum  
CC (potential).  
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.  
CC  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: A5003837; AAF59090.1; --  
CC FlyBase: FBgn0015034; Cyd4el.  
CC InterPro: IPR001128; Cytochrome\_P450.  
CC Pfam: PF00067; P450; 1.  
CC PRINTS: PR00385; P450.  
CC PROSITE: PS00086; CYTOCHROME P450; 1.  
CC Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;  
KW Endoplasmic reticulum; Hypothetical protein.  
KW 444 444 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).  
FT METAL 444 444  
SQ SEQUENCE 531 AA; 61086 MW; 42D45A86B5DB905 CRC64;  
Query Match 12.9%; Score 398.5; DB 1; Length 531;  
Best Local Similarity 25.2%; Pred. No. 2e-18;  
Matches 143; Conservative 91; Mismatches 217; Indels 131; Gaps 24;  
QY 26 YTALLVLVLSW-----ILVQWSLRKQKGRSFWIGATVEQLRN-----YHRMHW---- 72  
DB 8 FLALPLFLVTFELGLLRKRLMKLPQGPSMLPLVG-NAHQMGNTPEILNFFGWNHWEY 66  
QY 73 -----LVGYLSRHRVTVDMPFTSYTIVADPNVVEHLKTNFTNPKGIVYRSYMDVL 125  
DB 67 GKDNFRYWTGYISN-----IMVTNPKYMEFILSSQ-TLISKSDVY-DLTHFW 111  
QY 126 LGDGIENADGELWRKQRTASFEFASKNLDRDFAIVFREYSLKSLGILSQASKAGKVDM 185  
DB 112 LGLGLLTSTGSKWKKRKMTPAFHENILQDPHE-VNENSTKFIQDLKKVADGGNIEF 170  
QY 186 QELYMRMTLSDICKVGFVEIGTLSPDLSPNSFAQAFDAANIITLRFIDPLMRKRFH 245  
DB 171 QEEAHVLTLDVICDTANGVSNAM--ENRSSVVQAFKDIITYIKRAFSPWKRNYLFH 228  
QY 246 VGSE-ALLAQSIKLVDFTYSVIRRRKAEIVVEVRASQKQKMKHDILSR----FIELGEA 300  
DB 229 FAFEPYESTKLTQLDFTNEIIAKR-----IEVRKSGLEVIGIKADFSRKKMAFLDTLS 284  
QY 301 GDGGGFGDDKSLRDVVLNFIAGRTDTATTLGWFTHMASHPDVAFKRLREICAFEAER 360  
DB 285 SKVDGRPLTSQELVEEVSTWFEFGHDTTSGVGFVAVLLSRHPDEQELFNEQC----DV 340  
QY 361 AREGVTILVCGADADDKFAARVAQFAGLIIYDSLGKLVYLHACVTEITRLYPVAP-- 418  
DB 341 MGASGL-----GRDA-----TFQEIISTMKHLDLFTKEAQRIYSPVFFI 378  
QY 419 ---QDPKGILEDVLPDGTKVRAGGMVTVYVPSNGRMEYN--WGPDAASFRPERWINEG 473  
DB 379 GRFTEKDYVIDGDIIVPKGTLINLGLIM-----LGYNDRVFKDPHKQPERFDE-- 427  
QY 474 AFRNASPFKFTAFQAGPRICLGKDSAYLQMKMALAILFRFYSFELL----- 519  
DB 428 ---KPGFPEYVFPFAGPRNCIGQKFALEIKTVVSKLIIR--NFEVLFDALDELVSKOGYIS 482  
QY 520 -----BGH-----PVQYRMWMTLSMAHGLKVRV-SRAVC 547

DB 483 TTLGLQPAEKSRDAHNHXYDFILSASMTLKS-ENGLHLMKQRLVC 528  
RESULT 12  
CP5M DEBHA  
ID CP5M DEBHA STANDARD; PRT; 519 AA.  
AC Q9Y758;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Cytochrome P450 52A13 (BC 1.14.14.-) (Alkane hydroxylase 2) (Alkane-  
DE inducible p450alk 2) (DH-ALK2).  
DE CYP52A13 OR ALK2.  
GN Debaryomyces hansenii (Yeast) (*Torulasporea hansenii*).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Debaryomycetes.  
OX NCBI\_TaxID=4959;  
RN [1]\_TaxID=4959;  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 20317; PubMed=9931473;  
RX MEDLINE=99132287; Yada J.S., Loper J.C.;  
RA "Multiple p450alk (cytochrome P450 alkane hydroxylase) genes from the  
RT halotolerant yeast *Debaryomyces hansenii*.";  
RT Gene 226:139-146(1999).  
RL  
CC -!- FUNCTION: Together with an NADPH cytochrome P450 the enzyme system  
CC catalyzes the terminal hydroxylation as the first step in the  
CC assimilation of alkanes and fatty acids.  
CC -!- CATALYTIC ACTIVITY: Hydroxylation of N-alkanes at the terminal  
CC position.  
CC -!- INDUCTION: By N-alkanes.  
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.  
CC  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: AF103949; RAD22537.1; --  
CC HSP; P14779; IJPZ.  
CC InterPro: IPR001128; Cytochrome\_P450.  
CC Pfam: PF00067; P450; 1.  
CC PRINTS: PR00385; P450.  
CC PROSITE: PS00086; CYTOCHROME P450; 1.  
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme.  
FT METAL 466 466 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).  
SQ SEQUENCE 519 AA; 59607 MW; 1CE7D1EE23CEB5A CRC64;  
Query Match 12.8%; Score 395; DB 1; Length 519;  
Best Local Similarity 25.1%; Pred. No. 3.2e-18;  
Matches 141; Conservative 96; Mismatches 222; Indels 102; Gaps 17;  
QY 26 YTALLVLVLSWILVQWSLRKQKGR-----SWPVIGATVEQLRNHRMHD 71  
DB 16 YVVISALLVSFFIAHKISVARFKATHNCAASPEYKVNWFSLPLLYRLI-QVREGRLLD 74  
QY 72 WLVGYSLRHRTVTDMPFTSYTIV--ADPNVVEHLKTNFTNPKGIVYRSYMDVLGDD 129  
DB 75 FQAKYDVKALTFVIKVGPFVITRDENKAVLATQFNDPALG-TRAHFKPLGDD 133  
QY 130 IFNADGELWRKQRTASFEFASKNLDRDFAIVFREYSLKSLGILSQASKAGKVDMQELY 189  
DB 134 IFTLDGNKQKSRMLRPQFSREQVAHQAL--EPHLQELAKHRLAD--GETINIQDLF 189  
QY 190 MRMTLDSICKVGFVEIGTLSP-----DLP-----ENSPAQAFDAANIITLRFIDPLMR 239  
DB 190 FKLTVDVTFATEFLFGQSVGLKDAANDPPTDFGRSSFSANSFNTACTYGLTRAY----- 244  
QY 240 IKKFFHVGSEALLAQSIKLVDFTYSVIRRRKAEIVVEVRASQKQKMKHDILSRFIELGE 299

Matches 131; Conservative 74; Mismatches 183; Indels 88; Gaps 17;

QY 97 DPVNEVHLKTNFTNPKGIVVRSYMDVLGCGIFNADGELWKKQRTAFSEFASKNLRD 156  
 Db 101 DPNIKAMLATQENDFTLG-QRLSYFAPLIGKIGFTLDGEGNKHSLRMLRPPQSDOVGH 159  
 QY 157 FSAI-----VFYSYSLKLSGILLQSKAGKVVDMQELYMWTLDLSICKVGFVGEIGTLP 211  
 Db 160 VKMLEPHFQLLKXHIK-----NKGFFDIQELFFFTVDSATEFLFGSVSLKD 210  
 QY 212 D-----LPENSAFAQAFDAANIIITLRIDFLWRIKFRPHVSEALLAQSIKLV 260  
 Db 211 ESTGYDOEEDFAGKDFAEAFNKSQVYLSTRSL-----LQLLYLWLVNSDFKRCNKIVH 265  
 QY 261 EFTYSVIRRKAEIVVRASGKQKXKHDLSLFIELGEAGDDGGGDKSLRDVVLNF 320  
 Db 266 KFSYVIKK-----ALTATPELEKH-----SSYIFLYELAKQ-----TRDPVLRDQSLNI 312  
 QY 321 VIAGRDTTATLSWFTWMSHPDVAEKLRLRCAFEARERBEGVTLVLCG---GADAD 377  
 Db 313 LLAGRDTTAGLLSPAVFELGRNPEVWSKLREI-----GDKFGLDPD 354  
 QY 378 DKAFARVAQFAGLLTYDSLGLKLYLHACVTETLRLYPVQDPKGLIEDDVLPGD--- 433  
 Db 355 -----SRIED-----ISPELLKLCYLKAVINETLRLYPSVPRNGRFAAANTTLPHGGGPD 405  
 QY 434 ----TKVRAGMVTYVYVSMGRMEYNGWPDAAASFRPERKWNEDGAFNAPSPFKETAQAG 489  
 Db 406 GMSPIVRKGTVMYSYVALQRDEKYKQDANERPERWEPE---VRKLG-WAFLPFGG 462  
 QY 490 PRICLGKDSAYLQMKALAILFR-FYSFRLLGHPVQVYMMTILSMA--HGLKVRV 542  
 Db 463 PRICLQGFALTEASYVLVRLIQSFTEILESPDAPYPPAKLTLTMCFLDQAPVRI 518

RESULT 14  
 CP5K CANMA STANDARD; PRT; 524 AA.

AC Q9HCS2; Q9HCS1;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Cytochrome P450 4F12 (EC 1.14.14.1) (CYP4F12).  
 GN Homo sapiens (Human).  
 OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND CHARACTERIZATION.  
 RC TISSUE=Liver;  
 RX MEDLINE=21092695; PubMed=11162607;  
 RA Bylund J., Bylund M., Olin E.H.;  
 RT "cDNA cloning and expression of CYP4F12, a novel human cytochrome  
 P450.";  
 RL Biochem. Biophys. Res. Commun. 280:892-897(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A., TISSUE SPECIFICITY, CHARACTERIZATION, AND  
 RC TISSUE=Small intestine;  
 RX MEDLINE=21092733; PubMed=11162645;  
 RA Hashizume T., Imaoka S., Hiroi T., Terauchi Y., Fujii T., Miyazaki H.,  
 RT "cDNA cloning and expression of a novel cytochrome P450 (cyp4f12)  
 from human small intestine.";  
 CC -!- FUNCTION: Catalyzes leukotriene B(4) omega-hydroxylation and  
 arachidonic acid omega-hydroxylation but with an activity much  
 lower than that of CYP4F2. Catalyzes the hydroxylation of the  
 antihistamine ebastine.  
 CC -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +  
 oxidized flavoprotein + H(2)O.  
 CC -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum (By

Db 245 LQMFYFVNNDFKCKQVHDETFYVQKGLDMTP-----ELEK 285  
 QY 300 AGDDGGG-----GDDKSLRDVNLVFIAGBDTTATLSWFTWMSHPDVAEKLRR 351  
 Db 286 KSENGYVFLVLRQTRDPKVLQDQLNILLAGDRTTAGLLSFTFFELARHPRVFNKLKE 345  
 QY 352 ELCAFAEAREEGVTLVLCGGADADKAFARVAQFAGLLTYDSLGLKLYLHACVTETL 411  
 Db 346 EI-----YFAFGKGDARVSE-----ITPESLKKCEYLKWWNEML 381  
 QY 412 RLYPAVQDPKGLIEDDVL-----PDGFK-----VRAGMVTYVYVSMGRMEYNGWPDAA 463  
 Db 382 RLYPSVPNPRVAVKRTTLPRGGGPDGNSPIYVGKGTTVAYSVSTRMEYVYKGDAD 441  
 QY 464 RPERWINEGAFRNASPKFTAFQAGPRICLGKDSAYLQMK-MALAILFRYSRLLLEGH 522  
 Db 442 KPERW-----AESRKLGWYVFPNGPRICLQGFALTEASYIVTRLLQMPDKLELHDDR 496  
 QY 523 PVQYRMMTILSMHGLKVRVS 543  
 Db 497 PYPPAKSVHLTHCHQDGVVS 517

RESULT 13  
 CP5K CANMA STANDARD; PRT; 519 AA.

AC Q12589;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Cytochrome P450 52A11 (EC 1.14.14.-) (CYP52A11) (Alkane-inducible  
 P450-ALKs).  
 GN CYP52A11  
 OS Candida maltosa (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Mitosporic Saccharomycetales; Candida.  
 OX NCBI\_TaxID=5479;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95162975; PubMed=7865134;  
 RA Okuma M., Muraoka S., Tanimoto T., Fujii M., Ohta A., Takagi M.;  
 RT "CYP52 (cytochrome P450alk) multigene family in Candida maltosa:  
 identification and characterization of eight members.";  
 RL DNA Cell Biol. 14:163-173(1995).  
 CC -!- FUNCTION: Together with an NADPH cytochrome P450 the enzyme system  
 catalyzes the terminal hydroxylation as the first step in the  
 assimilation of alkanes and fatty acids.  
 CC -!- CATALYTIC ACTIVITY: Hydroxylation of N-alkanes at the terminal  
 position.  
 CC -!- INDUCTION: By N-alkanes.  
 CC -!- SIMILARITY: Belongs to the cytochrome P450 family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; D12719; BAA02214.1; -;  
 CC PIR; J50726; J50726.  
 CC HSP; P14779; 1J3P.  
 CC InterPro; IPR001128; Cytochrome\_P450.  
 CC Pfam; PF00067; P450; 1.  
 CC PRINTS; PR00385; P450.  
 CC PROSITE; PS00086; CYTOCHROME\_P450; 1.  
 CC Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme.  
 FT METAL 466 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).  
 SQ SEQUENCE 519 AA; 59476 MW; 70A9430ADEE38096 CRC64;

Query Match 12.6%; Score 387; DB 1; Length 519;  
 Best Local Similarity 27.5%; Pred. No. 1.1e-17;

```

456 WGPDAASFRPERWINEEDGAFRNASPFKTAFOAGPRICIGKDSAYLQMKXVALILRRFYS 515
434 W-PDVEYDPPFRDPENS--KGRSPLAFIPFGAGPRNCIGQAFAMEMKVVLALM--LLH 498
516 FELLEGHFPQYRMWMTILSMAHGLKVRVSRVAVCHGDLMDIVPIN 559
489 FRFLPDHTEPRKLELIMEA-----EGGLWLRVPELIN 520

RESULT 15
CP5D CANVA
ID CP5D CANVA STANDARD; PRT; 538 AA.
AC F16141;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cytochrome P450 52A4 (EC 1.14.14.-) (CYPLI1A4) (Alkane-inducible P450-
DE ALK3-A) (P450-CM2).
DE CYP52A4.
GN CN
OS Candida maltosa (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OC NCBI_TaxID=5479;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EH15D.
RX MEDLINE=89286595; PubMed=2735924;
RA Schunk W.-H., Kaergel E., Gross B., Wiedmann B., Mauerberger S.,
RA Koepke K., Kiessling U., Strauss M., Gaestel M., Mueller H.-G.;
RT "Molecular cloning and characterization of the primary structure of
RT the alkane hydroxylating cytochrome P-450 from the yeast Candida
RT maltosa".
RL Biochem. Biophys. Res. Commun. 161:843-850(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91229697; PubMed=2039569;
RA Ohkuma M., Tanimoto T., Yano K., Takagi M.;
RT "CYP52 (cytochrome P450alk) multigene family in Candida maltosa:
RT molecular cloning and nucleotide sequence of the two tandemly
RT arranged genes."
RL DNA Cell Biol. 10:271-282(1991).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=96311366; PubMed=8713123;
RA Zimmer T., Ohkuma M., Ohta A., Takagi M., Schunck W.H.;
RT "The CYP52 multigene family of Candida maltosa encodes functionally
RT diverse n-alkane-inducible cytochromes P450."
RL Biochem. Biophys. Res. Commun. 224:784-789(1996).
CC -!- FUNCTION: Together with an NADPH cytochrome P450 the enzyme system
CC catalyzes the terminal hydroxylation as the first step in the
CC assimilation of alkanes and fatty acids.
CC -!- CATALYTIC ACTIVITY: Hydroxylation of N-alkanes at the terminal
CC position.
CC -!- INDUCTION: By alkanes.
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement
or send an email to license@isb-sib.ch).

EMBL; X51932; CAA36198.1; -
DR EIR; S08668; O4CKA4.
DR HSSP; F14779; LJPZ.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
DR Electron transport; Oxidoreductase; Monooxygenase; Heme;
KW

```

Search completed: March 2, 2004, 15:28:12  
Job time : 19 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2004, 15:29:15 ; Search time 38 Seconds  
(without alignments)  
3267.319 Million cell updates/sec

Title: US-10-021-657-2

Perfect score: 3079  
Sequence: 1 EFGTREAHLTPATPSPFFPL.....ICMHACKGRWVSLVAVLKP 588

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 21153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3079	100.0	588	13	US-10-021-657-2
2	3079	100.0	588	14	US-10-412-000-2
3	770	25.0	158	13	US-10-021-657-4
4	770	25.0	158	14	US-10-412-000-4
5	401	13.0	492	15	US-10-369-493-3749
6	396	12.9	517	9	US-09-911-781-32
7	396	12.9	517	10	US-09-976-800-100
8	396	12.9	517	14	US-10-138-838-100
9	396	12.9	517	14	US-10-139-031-100
10	396	12.9	517	14	US-10-138-905-100
11	396	12.9	517	14	US-10-138-916-100
12	396	12.9	517	14	US-10-139-296-100
13	396	12.9	517	14	US-10-139-218-100
14	396	12.9	517	14	US-10-400-902-32
15	396	12.9	517	14	US-10-403-660-100

16	396	12.9	517	15	US-10-138-898-100	Sequence 100, App
17	395	12.8	517	10	US-09-976-800-101	Sequence 101, App
18	395	12.8	517	14	US-10-138-838-101	Sequence 101, App
19	395	12.8	517	14	US-10-139-031-101	Sequence 101, App
20	395	12.8	517	14	US-10-138-905-101	Sequence 101, App
21	395	12.8	517	14	US-10-138-916-101	Sequence 101, App
22	395	12.8	517	14	US-10-139-296-101	Sequence 101, App
23	395	12.8	517	14	US-10-139-218-101	Sequence 101, App
24	395	12.8	517	14	US-10-403-660-101	Sequence 101, App
25	395	12.8	517	15	US-10-138-898-101	Sequence 101, App
26	393.5	12.8	546	15	US-10-291-265-405	Sequence 405, App
27	393	12.8	522	10	US-09-976-800-97	Sequence 97, Appl
28	393	12.8	522	14	US-10-138-838-97	Sequence 97, Appl
29	393	12.8	522	14	US-10-139-031-97	Sequence 97, Appl
30	393	12.8	522	14	US-10-138-905-97	Sequence 97, Appl
31	393	12.8	522	14	US-10-139-916-97	Sequence 97, Appl
32	393	12.8	522	14	US-10-139-296-97	Sequence 97, Appl
33	393	12.8	522	14	US-10-139-218-97	Sequence 97, Appl
34	393	12.8	522	14	US-10-403-660-97	Sequence 97, Appl
35	393	12.8	522	15	US-10-138-898-97	Sequence 97, Appl
36	389	12.6	517	10	US-09-976-800-112	Sequence 112, App
37	389	12.6	517	14	US-10-138-838-112	Sequence 112, App
38	389	12.6	517	14	US-10-139-031-112	Sequence 112, App
39	389	12.6	517	14	US-10-138-905-112	Sequence 112, App
40	389	12.6	517	14	US-10-138-916-112	Sequence 112, App
41	389	12.6	517	14	US-10-139-296-112	Sequence 112, App
42	389	12.6	517	14	US-10-139-218-112	Sequence 112, App
43	389	12.6	517	14	US-10-403-660-112	Sequence 112, App
44	389	12.6	517	15	US-10-138-898-112	Sequence 112, App
45	389	12.6	523	10	US-09-976-800-95	Sequence 95, Appl

#### ALIGNMENTS

#### RESULT 1

US-10-021-657-2  
; Sequence 2, Application US/10021657  
; Publication No. US20020083483A1  
; GENERAL INFORMATION:  
; APPLICANT: ALBERTSEN, MARC C.  
; APPLICANT: FOX, TIM  
; APPLICANT: HUFEMAN, GARY  
; APPLICANT: TRIMMELL, MARY  
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES MEDIATING MALE FERTILITY AND  
; TITLE OF INVENTION: METHOD OF USING SAME  
; FILE REFERENCE: 1148  
; CURRENT APPLICATION NUMBER: US/10/021,657  
; CURRENT FILING DATE: 2001-12-14  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 588  
; TYPE: PRT  
; ORGANISM: Zea mays  
US-10-021-657-2

Query Match	100.0%;	Score	3079;	DB	13;	Length	588;
Best Local Similarity	100.0%;	Pred. No.	1.2e-313;				
Matches	588;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
Qy	1	EFGTREAHLTPATPSPFFPLAGPHKYIALLVLSWILVQWSLRKQKGRSPWPVIGATV	60				
Db	1	EFGTREAHLTPATPSPFFPLAGPHKYIALLVLSWILVQWSLRKQKGRSPWPVIGATV	60				
Qy	61	BQLKNYHRMHDWLGVLSRHRVTVDMPFTSYTYIADPVNVHVLKTNFTNPKGIVYRS	120				
Db	61	BQLKNYHRMHDWLGVLSRHRVTVDMPFTSYTYIADPVNVHVLKTNFTNPKGIVYRS	120				
Qy	121	YMDVLLGDGIFNADGELWKRKORTAFEFASKNLRDPSAIVFREYSLSLGSILQSKAG	180				
Db	121	YMDVLLGDGIFNADGELWKRKORTAFEFASKNLRDPSAIVFREYSLSLGSILQSKAG	180				

QY 181 KVVDMQELYNRMWLDLSCKVGFVGEIGTSLSPDLPENSFAQAFDAANIIITLRFIDPLWRI 240  
 Db 181 KVVDMQELYNRMWLDLSCKVGFVGEIGTSLSPDLPENSFAQAFDAANIIITLRFIDPLWRI 240  
 QY 241 KRFFHVGSEALLAQSIKLVDEFTYSVIRRRKAEIIVEVRASGKQKMKHDILSRFIELGEA 300  
 Db 241 KRFFHVGSEALLAQSIKLVDEFTYSVIRRRKAEIIVEVRASGKQKMKHDILSRFIELGEA 300  
 QY 301 GDDGGGFGDDKSLRDVVNFVIAGRDITATTLSWFTHMAMSHPDVAEKLRRRELCAFEAE 360  
 Db 301 GDDGGGFGDDKSLRDVVNFVIAGRDITATTLSWFTHMAMSHPDVAEKLRRRELCAFEAE 360  
 QY 361 AREEGVTLVLCGGADADDKAFARVAQFAGLLTYDSLGKLVYLHACVTETLRLYPVPOD 420  
 Db 361 AREEGVTLVLCGGADADDKAFARVAQFAGLLTYDSLGKLVYLHACVTETLRLYPVPOD 420  
 QY 421 PKGILEDDVLPDGTGKVRAGGWTVPYSGRMEYNNWGPDAASFRPERWINEGAFRNASP 480  
 Db 421 PKGILEDDVLPDGTGKVRAGGWTVPYSGRMEYNNWGPDAASFRPERWINEGAFRNASP 480  
 QY 481 FKFTAFQAGPRICLGKDSAYLQMKMALAILFRFYSFRLLEGHVPQYRMWMTILSMAHGLKV 540  
 Db 481 FKFTAFQAGPRICLGKDSAYLQMKMALAILFRFYSFRLLEGHVPQYRMWMTILSMAHGLKV 540  
 QY 541 RVSRVCHGDLMDIVPLNPRQITVLQICMHACKGRWVSLVAVLKP 588  
 Db 541 RVSRVCHGDLMDIVPLNPRQITVLQICMHACKGRWVSLVAVLKP 588

RESULT 2  
 US-10-412-000-2  
 ; Sequence 2, Application US/10412000  
 ; Publication No. US20030182689A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ALBERTSEN, MARC C.  
 ; APPLICANT: FOX, TIM  
 ; APPLICANT: HUFFMAN, GARY  
 ; APPLICANT: TRIMMELL, MARY  
 ; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES MEDIATING MALE FERTILITY AND  
 ; FILE REFERENCE: 1148  
 ; CURRENT APPLICATION NUMBER: US/10/412,000  
 ; PRIOR FILING DATE: 2003-04-11  
 ; PRIOR FILING DATE: 2001-06-11  
 ; NUMBER OF SEQ ID NOS: 7  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 2  
 ; LENGTH: 588  
 ; TYPE: PRT  
 ; ORGANISM: Zea mays  
 ; OTHER INFORMATION: "Xaa" may be any, other or unknown amino acid  
 US-10-412-000-2

Query Match 100.0%; Score 3079; DB 14; Length 588;  
 Best Local Similarity 100.0%; Pred. No. 1-2e-313;  
 Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 EFGTREAHLTPATSPFPPLAGPHKYIALLVLSWILVQWRSIRKQKGRSPWPVIGATV 60  
 Db 1 EFGTREAHLTPATSPFPPLAGPHKYIALLVLSWILVQWRSIRKQKGRSPWPVIGATV 60  
 QY 61 EQLRNHRMHDWLVGYLSRHRVTVDMPFTSYTIVADPVNVEHLKTNFTNPKGIYVRS 120  
 Db 61 EQLRNHRMHDWLVGYLSRHRVTVDMPFTSYTIVADPVNVEHLKTNFTNPKGIYVRS 120  
 QY 121 YMDVLLGDGIFNADGELWRKORTASPEFASKNLRDFAISVFREYSKLGILSQASKAG 180  
 Db 121 YMDVLLGDGIFNADGELWRKORTASPEFASKNLRDFAISVFREYSKLGILSQASKAG 180  
 QY 181 KVVDMQELYNRMWLDLSCKVGFVGEIGTSLSPDLPENSFAQAFDAANIIITLRFIDPLWRI 240  
 Db 181 KVVDMQELYNRMWLDLSCKVGFVGEIGTSLSPDLPENSFAQAFDAANIIITLRFIDPLWRI 240

QY 241 KRFFHVGSEALLAQSIKLVDEFTYSVIRRRKAEIIVEVRASGKQKMKHDILSRFIELGEA 300  
 Db 241 KRFFHVGSEALLAQSIKLVDEFTYSVIRRRKAEIIVEVRASGKQKMKHDILSRFIELGEA 300  
 QY 301 GDDGGGFGDDKSLRDVVNFVIAGRDITATTLSWFTHMAMSHPDVAEKLRRRELCAFEAE 360  
 Db 301 GDDGGGFGDDKSLRDVVNFVIAGRDITATTLSWFTHMAMSHPDVAEKLRRRELCAFEAE 360  
 QY 361 AREEGVTLVLCGGADADDKAFARVAQFAGLLTYDSLGKLVYLHACVTETLRLYPVPOD 420  
 Db 361 AREEGVTLVLCGGADADDKAFARVAQFAGLLTYDSLGKLVYLHACVTETLRLYPVPOD 420  
 QY 421 PKGILEDDVLPDGTGKVRAGGWTVPYSGRMEYNNWGPDAASFRPERWINEGAFRNASP 480  
 Db 421 PKGILEDDVLPDGTGKVRAGGWTVPYSGRMEYNNWGPDAASFRPERWINEGAFRNASP 480  
 QY 481 FKFTAFQAGPRICLGKDSAYLQMKMALAILFRFYSFRLLEGHVPQYRMWMTILSMAHGLKV 540  
 Db 481 FKFTAFQAGPRICLGKDSAYLQMKMALAILFRFYSFRLLEGHVPQYRMWMTILSMAHGLKV 540  
 QY 541 RVSRVCHGDLMDIVPLNPRQITVLQICMHACKGRWVSLVAVLKP 588  
 Db 541 RVSRVCHGDLMDIVPLNPRQITVLQICMHACKGRWVSLVAVLKP 588

RESULT 3  
 US-10-021-657-4  
 ; Sequence 4, Application US/10021657  
 ; Publication No. US20020083483A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ALBERTSEN, MARC C.  
 ; APPLICANT: FOX, TIM  
 ; APPLICANT: HUFFMAN, GARY  
 ; APPLICANT: TRIMMELL, MARY  
 ; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES MEDIATING MALE FERTILITY AND  
 ; FILE REFERENCE: 1148  
 ; CURRENT APPLICATION NUMBER: US/10/021,657  
 ; CURRENT FILING DATE: 2001-12-14  
 ; NUMBER OF SEQ ID NOS: 7  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 4  
 ; LENGTH: 158  
 ; TYPE: PRT  
 ; ORGANISM: Sorghum sp.  
 ; FEATURE:  
 ; NAME/KEY: MOD RES  
 ; LOCATION: (1)-(158)  
 ; OTHER INFORMATION: "Xaa" may be any, other or unknown amino acid  
 US-10-021-657-4

Query Match 25.0%; Score 770; DB 13; Length 158;  
 Best Local Similarity 94.3%; Pred. No. 3e-72;  
 Matches 149; Conservative 3; Mismatches 6; Indels 0; Gaps 0;  
 QY 87 MFPTSYTYIADPVNVEHLKTNFTNPKGIYVRSYNDVLLGDGIFNADGELWRKORTAS 146  
 Db 1 MFPTSYTYIADPVNVEHLKTNFTNPKGIYVRSYNDVLLGDGIFNADGELWRKORTAS 60  
 QY 147 FEFASKNLRDFAISVFREYSKLGILSQASKAGKVVDMQELYMRTLDSICXVGFVGI 206  
 Db 61 FEFASKNLRDFAISVFREYSKLGILSQASKAGKVVDMQELYMRTLDSICXVGFVGI 120  
 QY 207 GTLSPLDLPENSFAQAFDAANIIITLRFIDPLWRIKFF 244  
 Db 121 GTLSPLDLPENSFAQAFDAANIIITLRFIDPLWRIKFF 158  
 RESULT 4  
 US-10-412-000-4  
 ; Sequence 4, Application US/10412000  
 ; Publication No. US20030182689A1  
 ; GENERAL INFORMATION:

APPLICANT: ALBERTSEN, MARC C.  
APPLICANT: FOX, TIM  
APPLICANT: HUFFMAN, GARY  
APPLICANT: TRIMMELL, MARY  
TITLE OF INVENTION: NUCLEOTIDE SEQUENCES MEDIATING MALE FERTILITY AND  
TITLE OF INVENTION: METHOD OF USING SAME  
FILE REFERENCE: 1148  
CURRENT APPLICATION NUMBER: US/10/412,000  
CURRENT FILING DATE: 2003-04-11  
PRIOR APPLICATION NUMBER: US/09/670,153  
PRIOR FILING DATE: 2001-06-11  
NUMBER OF SEQ ID NOS: 7  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 4  
LENGTH: 158  
TYPE: PRT  
ORGANISM: Sorghum sp.  
FEATURE:  
NAME/KEY: MOD RES  
LOCATION: (1)-(158)  
OTHER INFORMATION: "Xaa" may be any, other or unknown amino acid  
US-10-412-000-4

Query Match 25.0%; Score 770; DB 14; Length 158;  
Best Local Similarity 94.3%; Pred. No. 3e-72;  
Matches 149; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 87 MPFTSYTIADPVNVEHLKNTNPKGIVYRSDVLLGDFINADGELWRKORTAS 146  
DB 1 MPFTSYTIADPVNVEHLKNTNPKGIVYRSDVLLGDFINADGELWRKORTAS 60  
QY 147 FEFASKNLRDSAIYFREVSKLSGILSOAKAGKVVDMQELYMWTLDSCYKGFYEI 206  
DB 61 FEFASKNLRDSANVFRYSKLSGILSOAKAGKVVDMQELYMWTLDSCYKGFYEI 120  
QY 207 GTLSPLDLPNSPAQAFDAANIITLRFIDPLWRKRF 244  
DB 121 GTLSPLDLPNSFXQAFDAANIITLRFHPLWRKRF 158

## RESULT 5

US-10-369-493-3749  
Sequence 3749, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:

APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 3749  
LENGTH: 492  
TYPE: PRT  
ORGANISM: Neurospora crassa  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1)..(492)  
OTHER INFORMATION: unsure at all Xaa locations  
US-10-369-493-3749

## Query Match

Best Local Similarity 13.0%; Score 401; DB 15; Length 492;  
Matches 153; Conservative 71; Mismatches 177; Indels 162; Gaps 26;

QY 66

YHRMHDLVGLSRRTVTVDMPFTSYTIADPVNVEHLKNTNPKGIVYRSDVLL 125

DB 1 WQMFSGGTYGTVLGHRIITS-----EPENKAILATQEDYKGGFGHWEKDF 55  
QY 126 LGDGFNADGELWRKORTASFEFASKNLRDFA-----IVFREYSL--KLSG---ILS 174  
DB 56 LGDSIFITDGLDWHASQLIRPQFIKRVSDLOCFENHMQMLFRTIANGGALNGEDQWVD 115  
QY 175 QASKAGKVVDMQELYMWTLDSCYKGFYEIYGLTSPDL-----ENS-----F 218  
DB 116 MEAGNGKPVDSIDLFYRYTLDAATDFLLGDKDIKSLRFLVPNASPPICXNATENSTPVQPF 175  
QY 219 AQAQF-DAANIITLRFIDPLWRI--KRFEHVSEALLAQSIKLVD-FTYSVIRRRKAEI 274  
DB 176 ADAFOEVQVQIVTARAGPLNRFVPKTFWEG-----LKVIDETINFIIDRALRDE 227  
QY 275 VEVRASGQKQKMDHILSRFIELGEAGDGG-----GGFGDDKS-LRDVVNLFVIAGR 326  
DB 228 BELASKSK-----GDEGYTFELHALAGFTKNRQVLHDQLMAVLLAGRD 269  
QY 327 TTATILSKFWTHMAMSHDPVAEKLRRBELCAFEARAREEGVTILVLCGGADADDKAFARVA 386  
DB 270 TTACTLGSAIYELARHPEAVAKLRAEI-----LSVVG-----PDRA----- 305  
QY 387 QFAGLLTYDSLGLVYLHACVTEITLRYPAVP--QDP-----KGL 425  
DB 306 -----PYDDLKSKMILQNVNTEITLRYLPVVPFQVETSFSTRSCFLTFIVISVRAL 360  
QY 426 EDDVLPDGTKVRAAG-----MVTYVPYSMGEMEYNWG-----PDAASFRPE 466  
DB 361 KDTTLP-----RGGGPDGSGQPIVILKDTVPVGYSPLAMQRRPDLVPVSEKFPDVMESPD 415  
QY 467 RWINEDGAFRNASPFKTAFOAGPRICLGKDSAYLQMKMALAILFRFYSFELLE----- 520  
DB 416 RWFH-----WQPKWQXIPFNGSGPRICIGQOQFALTENGIVLTRLFQRYD-RWVSYMD 469  
QY 521 -GHPVQVQRMVT--ILSMAGHLKV 540  
DB 470 GSKP---RMKTDIVLMPGDGVKV 489

## RESULT 6

US-09-911-781-32  
Sequence 32, Application US/09911781  
Patent No. US20020034788A1  
GENERAL INFORMATION:

APPLICANT: Craft, David L.  
APPLICANT: Wilson, C. Ron  
APPLICANT: Eirich, Dudley  
APPLICANT: Zhang, Yeyan  
TITLE OF INVENTION: USE OF CYP52A2A PROMOTER TO INCREASE GENE EXPRESSION IN YEAST  
FILE REFERENCE: U0012 OS/OAAP (1010-49)  
CURRENT APPLICATION NUMBER: US/09/911.781  
CURRENT FILING DATE: 2001-07-24  
NUMBER OF SEQ ID NOS: 34  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 32  
LENGTH: 517  
TYPE: PRT  
ORGANISM: Candida tropicalis  
US-09-911-781-32

## Query Match 12.9%; Score 396; DB 9; Length 517;

Best Local Similarity 26.6%; Pred. No. 3.5e-32;  
Matches 139; Conservative 78; Mismatches 162; Indels 144; Gaps 19;

QY 93 TYIA-----DPNVNVEHLKNTNPKGIVYRSDVLLGDFINADGELWRKQ 141

DB 84 TYVSLFGTRIVVTYKDPENIKAILATQFDFSLGKRHTLFXP-LLGDFITLDGEGWHS 142

QY 142 RKTASFFFAKKNLRDFAI-----VPREYSKLSGILSOAKAGKVVDMQELYMWTLD 196

DB 143 RALRPPQFAREQVAHVTSLEPHFOLLKKHLK-----HKGEYFDIQELFFRFTVDS 193

[illegible]

RESULT 7  
US-09-976-800-100  
; Sequence 100, Application US/09976800  
; Publication No. US2003007795A1  
; GENERAL INFORMATION:  
; APPLICANT: Wilson, Ron C.  
; APPLICANT: Craft, David L.  
; APPLICANT: Birch, Dudley  
; APPLICANT: Esboo, Mark  
; APPLICANT: Madduri, Krishna M.  
; APPLICANT: Cornett, Cathy A.  
; APPLICANT: Brenner, Alfred A.  
; APPLICANT: Tang, Maria  
; APPLICANT: Loper, John C.  
; APPLICANT: Gleeson, Martin  
; TITLE OF INVENTION: CYTOCHROME P450 MONOOXYGENASE AND NADPH CYTOCHROME P450 OXIDOREDUCTASE  
; TITLE OF INVENTION: GENES AND PROTEINS RELATED TO THE OMEGA HYDROXYLASE COMPLEX OF CLOSTRIDIUM THERMOPHILUM  
; TITLE OF INVENTION: TROPICALIS AND METHODS RELATING THERETO  
; FILE REFERENCE: 1010-16  
; CURRENT APPLICATION NUMBER: US/09/976,800  
; CURRENT FILING DATE: 2001-10-12  
; NUMBER OF SEQ ID NOS: 118  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 100  
; LENGTH: 517  
; TYPE: PRT  
; ORGANISM: CANDIDATROPICALIS  
US-09-976-800-100

Qy	246	VGSALLAQSIKLYDET-YSVIRRRKAIEV-V-RAGS-----KOEKKHDLISRFI	295
Ddb	249	LNNKEFRDCTKLVKHKFTNYVQKALDASPELEKSGYVFLVELVKQR-----	298
Qy	296	ELGEAGDGGGFGDDKSLRDVVLNFVIAGRDTTATTLSWFTHMAMSHPDVAEKLRELCA	355
Ddb	299	-----DPNVLROQSLNILIAGRDTTAGLLSPAFAVELARHPETWAKLREEI--	343
Qy	356	FEARAREEGTVILCGADADDKAFAPRAVAQAQFAGLLTYDSLGLVYLHACVTETRLYP	415
Ddb	344	-----EQQFGL-----GED-----SRVEE---ITFESLKRCYELKPAFLNETLRIYP	381
Qy	416	AVPODPKGILEDDVLP-----DGTFK---VRAGGMVTVPYSMGRMENWNPGDAASRPFR	467
Ddb	382	SVPNFRIATKNNTLPRGGSDGTSPILIOKGAVSYGINSTHLDPVYVGDAAEFRPER	441
Qy	468	WINEDGAFRNASPPKFTAQAPPICIGKSAYLOMKVALAILFRFYSFLLLEGHPVOYR	527
Ddb	442	WFEPSS---TKKLGWAYLFPNGGPICIGQQOFA-----	470
Qy	528	MWTLSMAHGLKVRSRAVCHGDLDMDIV--PLNPROIITLVQLQ	568
Ddb	471	----LTEAGYVLRLVQEFESHVRLDPDEVPPKELTNLTCLQ	509

RESULT 8  
US-10-138-838-100  
; Sequence 100, Application US/10138838  
; Publication No. US20030049821A1  
; GENERAL INFORMATION:  
; APPLICANT: Wilson, Ron C.  
; APPLICANT: Craft, David L.  
; APPLICANT: Eirich, Dudley  
; APPLICANT: Eshoo, Mark  
; APPLICANT: Madduri, Krishna M.  
; APPLICANT: Cornett, Cathy A.  
; APPLICANT: Brenner, Alfred A.  
; APPLICANT: Tang, Maria  
; APPLICANT: Loper, John C.  
; APPLICANT: Gleeson, Martin  
; TITLE OF INVENTION: CYTOCHROME P450 MONOOXYGENASE AND NADPH CYTOCHROME  
; TITLE OF INVENTION: P450 OXIDOREDUCTASE  
; TITLE OF INVENTION: GENES AND PROTEINS RELATED TO THE OMEGA HYDROXYLASE COMPLEX OF  
; TITLE OF INVENTION: TROPICALIS AND METHODS RELATING THERETO  
; FILE REFERENCE: 1010-16  
; CURRENT APPLICATION NUMBER: US/10/138,838  
; CURRENT FILING DATE: 2002-05-03  
; PRIOR APPLICATION NUMBER: US/09/976,800  
; PRIOR FILING DATE: 2001-10-12  
; NUMBER OF SEQ ID NOS: 118  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 100  
; LENGTH: 517  
; TYPE: PR1  
; ORGANISM: CANDIDATROPICALIS  
US-10-138-838-100

Query Match	12.9%;	Score 396;	DB 14;	Length 517;
Best Local Similarity	26.6%;	Pred. No. 3.5e-32;		
Matches 139;	Conservative 78;	Mismatches 162;	Indels 144;	Gaps 19;
QY	93	TYIA-----DPVNVHVLKTNFTNYPKGIYVRSYMDVLGLDGFNADGELWRKQ	141	
DB	84	TVYSILFGTRIIVTKDPENIKAILATQGFDSLGKRHTLTKP-LLGDGIFTLLDGEWKHS	142	
QY	142	RKTASPEFAKSLRDPFSAL-----VFREYSKLISQASKAGKGVDMQELMYRMTLDS	196	
DB	143	RAMLAPQFAREQVAHVTSLEPHFOLLKXHLK-----HKGEYFDIQELFERTVDS	193	
QY	197	ICKVGFGEIGTLPSPD-----LPENSGAFAFDAANIITLRFIDPLMRKRFHH	245	
DB	194	ATEFFLFGESVHSIKDESGINODDIDFAGKQDFAESFNKAQBYLAITL-----VOTFFY	248	



194 ATEFTFGSVHSLKDESGINQDDIDFAGRKDFAESFKAQBYLAIRTL-----VOTFFW 248

246 VGSEALLAQSIKLVDEFT--YSVIRRRKAEIVEV--RASG-----KQEKMKHDLISRFI 295

249 LVNKEFRDCTKLVHKFTNYVOKALDASPELEKQSGVFLYELVKQTR----- 298

296 ELGEAGDGGGFGDKSLRDVNLVFIAGDRTTATLTSWFTMHAMSHDPDAEKLRELCA 355

299 -----DPNVLRDQSLNLLAGDRTTAGLLSFAVFELARHPFIWAKLRBEI-- 343

356 FEARAREEGVTVLVLCGGADADKFAARVAQFAGLLTYDSLGLKLVYLHACVTETRLYP 415

344 -----EQOFGU-----GED-----ITFESLKRCEYLKAFNEILRIYP 381

416 AVQDPKGILEDVLP-----DGTK-----VRAGGMVTVYPYSMGRMEYNWGPDAASFRPER 467

382 SVPRNFRIATKNTLPRGGSGDGTSPILIQGEAVSYGINSHTLDPVYVYGPDAAEFRPER 441

468 WINEDGAFRNAPFKFTAFQAGPRICLGKDSAYLQMKQALAILFRFYSFLLGHPVQYR 527

442 WFEPSS---TKLGMAYLPFNGGPRICLGQOFA----- 470

528 MMTILSMAHGLKVRVSRVAVCHGDLMDIV--PLNPRQITLVLQ 568

471 ----LTEAGYVLRLVQGFHSVRLDDEVPYPPKRLTNLTWCLQ 509

Db

RESULT 11

US-10-138-916-100

Sequence 100, Application US/10138916

Publication No. US20030073220A1

GENERAL INFORMATION:

APPLICANT: Wilson, Ron C.

APPLICANT: Craft, David L.

APPLICANT: Birch, Dudley

APPLICANT: Eshoo, Mark

APPLICANT: Madduri, Krishna M.

APPLICANT: Cornett, Cathy A.

APPLICANT: Brenner, Alfred A.

APPLICANT: Tang, Maria

APPLICANT: Loper, John C.

APPLICANT: Gleeson, Martin

TITLE OF INVENTION: CYTOCHROME P450 MONOOXYGENASE AND NADPH CYTOCHROME

TITLE OF INVENTION: GENES AND PROTEINS RELATED TO THE OMEGA HYDROXYLASE COMPLEX OF

TITLE OF INVENTION: TROPICALIS AND METHODS RELATING THERETO

FILE REFERENCE: 1010-16

CURRENT APPLICATION NUMBER: US/10/138,916

PRIOR FILING DATE: 2002-05-03

PRIOR FILING DATE: 2001-10-12

PRIOR FILING DATE: 1999-04-30

NUMBER OF SEQ ID NOS: 118

SOFTWARE: Patent in version 3.1

SEQ ID NO 100

LENGTH: 517

TYPE: PRT

ORGANISM: CANDIDATROPICALIS

US-10-138-916-100

Query Match 12.9%; Score 396; DB 14; Length 517;

Best Local Similarity 26.6%; Pred. No. 3.5e-32;

Matches 139; Conservative 78; Mismatches 162; Indels 144; Gaps 19;

93 TYIA-----DPNVNVEHLKTNFTNPKGIYVRSYMDVLGIGIFNADGELWRKQ 141

84 TYVSLFEGTRIVTKDPENIKAILATQGFDSLGKRLTFKP--LLGDGIFLDGEGWKHS 142

142 KTAFAEFASKNLRDFAI-----VFREYSKLGSILSQASKAGKVDVMDQELYWRMTLDS 196

143 RAMLRPQAREQVAHVTSLEPHFQLLKKHLK-----HKGEYFDIQELFFRFTVDS 193

197 ICKVGFGEIGTSLSPD-----LPENSAFAQADPAANIITLRFIDPLWRKIFPH 245

Db

194 ATEFTFGSVHSLKDESGINQDDIDFAGRKDFAESFKAQBYLAIRTL-----VOTFFW 248

246 VGSEALLAQSIKLVDEFT--YSVIRRRKAEIVEV--RASG-----KQEKMKHDLISRFI 295

249 LVNKEFRDCTKLVHKFTNYVOKALDASPELEKQSGVFLYELVKQTR----- 298

296 ELGEAGDGGGFGDKSLRDVNLVFIAGDRTTATLTSWFTMHAMSHDPDAEKLRELCA 355

299 -----DPNVLRDQSLNLLAGDRTTAGLLSFAVFELARHPFIWAKLRBEI-- 343

356 FEARAREEGVTVLVLCGGADADKFAARVAQFAGLLTYDSLGLKLVYLHACVTETRLYP 415

344 -----EQOFGU-----GED-----ITFESLKRCEYLKAFNEILRIYP 381

416 AVQDPKGILEDVLP-----DGTK-----VRAGGMVTVYPYSMGRMEYNWGPDAASFRPER 467

382 SVPRNFRIATKNTLPRGGSGDGTSPILIQGEAVSYGINSHTLDPVYVYGPDAAEFRPER 441

468 WINEDGAFRNAPFKFTAFQAGPRICLGKDSAYLQMKQALAILFRFYSFLLGHPVQYR 527

442 WFEPSS---TKLGMAYLPFNGGPRICLGQOFA----- 470

528 MMTILSMAHGLKVRVSRVAVCHGDLMDIV--PLNPRQITLVLQ 568

471 ----LTEAGYVLRLVQGFHSVRLDDEVPYPPKRLTNLTWCLQ 509

Db

RESULT 12

US-10-139-296-100

Sequence 100, Application US/10139296

Publication No. US20030148486A1

GENERAL INFORMATION:

APPLICANT: Wilson, Ron C.

APPLICANT: Craft, David L.

APPLICANT: Birch, Dudley

APPLICANT: Eshoo, Mark

APPLICANT: Madduri, Krishna M.

APPLICANT: Cornett, Cathy A.

APPLICANT: Brenner, Alfred A.

APPLICANT: Tang, Maria

APPLICANT: Loper, John C.

APPLICANT: Gleeson, Martin

TITLE OF INVENTION: CYTOCHROME P450 MONOOXYGENASE AND NADPH CYTOCHROME

TITLE OF INVENTION: P450 OXIDOREDUCTASE

TITLE OF INVENTION: GENES AND PROTEINS RELATED TO THE OMEGA HYDROXYLASE COMPLEX OF

TITLE OF INVENTION: TROPICALIS AND METHODS RELATING THERETO

FILE REFERENCE: 1010-16

CURRENT APPLICATION NUMBER: US/10/139,296

PRIOR FILING DATE: 2002-05-03

PRIOR FILING DATE: US/09/976,800

PRIOR FILING DATE: 2001-10-12

NUMBER OF SEQ ID NOS: 118

SOFTWARE: Patent in version 3.1

SEQ ID NO 100

LENGTH: 517

TYPE: PRT

ORGANISM: CANDIDATROPICALIS

US-10-139-296-100

Query Match 12.9%; Score 396; DB 14; Length 517;

Best Local Similarity 26.6%; Pred. No. 3.5e-32;

Matches 139; Conservative 78; Mismatches 162; Indels 144; Gaps 19;

93 TYIA-----DPNVNVEHLKTNFTNPKGIYVRSYMDVLGIGIFNADGELWRKQ 141

84 TYVSLFEGTRIVTKDPENIKAILATQGFDSLGKRLTFKP--LLGDGIFLDGEGWKHS 142

142 KTAFAEFASKNLRDFAI-----VFREYSKLGSILSQASKAGKVDVMDQELYWRMTLDS 196

143 RAMLRPQAREQVAHVTSLEPHFQLLKKHLK-----HKGEYFDIQELFFRFTVDS 193

197 ICKVGFGEIGTSLSPD-----LPENSAFAQADPAANIITLRFIDPLWRKIFPH 245

Db 194 ATEFLGESVHSLKDESIGINQDDIDFAGRKDFAESFNKAQEYLAIRTL-----VQTFYV 248  
QY 246 VGSEALLAQSILKLVDEFT--YSVIRRRKAEIIVEV--RASG-----KQKMKHDLISRFI 295  
Db 249 LVNNKEFRDCTKLHKFTNYVYQKALDASPEELEKQSGVVFYELVKQTR-----298  
QY 296 ELGEAGDDGGGDDKSLRDVNLVNFVIAGRDTTATTLTSMFTMAMSHPDVAEKLRELCA 355  
Db 299 -----DENVLRDQSLNILLAGRTDTTAGLLSFAVFLARHPETWAKLREI--343  
QY 356 FEARAREEGVTILVLCGGADADKAFARVAQAGLLTVDSLGKLVHLHACVTETLRLYP 415  
Db 344 -----EQQFGL-----GED-----SRVEE-----ITFSLKRCBYLKAFLNETLRIYP 381  
QY 416 AVPDQPKGILEDDVLP-----DGTK-----VRAGMVTYVPVYSGMGEYNNWGPDAASFRPER 467  
Db 382 SVERNFRIATKNTLPRGGSGDGTSPILQKGAVSYGINSHTLDPVYVGPDAAEFRPER 441  
QY 468 WINEDGAFRNASPFKFTAFQAGPRICLGKDSAYLQMKMALAILFRFYSFLLLEGHPVQYR 527  
Db 442 WPEPS---TKLGLWYLPFNGGPRICLGQOFA-----470  
QY 528 MMTILSMAHGLKVRVSRAVCHGDLMDIV--PLNPRQITILVLQ 568  
Db 471 ----LTEAGYVLRVLVQEFSHVLDPEVYPPKRLTNLTWCLQ 509

RESULT 13  
US-10-139-218-100  
; Sequence 100, Application US/10139218  
; Publication No. US20030153060A1  
; GENERAL INFORMATION:  
; APPLICANT: Wilson, Ron C.  
; APPLICANT: Craft, David L.  
; APPLICANT: Erich, Dudley  
; APPLICANT: Esboo, Mark  
; APPLICANT: Madhuri, Krishna M.  
; APPLICANT: Cornett, Cathy A.  
; APPLICANT: Brenner, Alfred A.  
; APPLICANT: Tang, Maria  
; APPLICANT: Loper, John C.  
; APPLICANT: Gleeson, Martin  
; TITLE OF INVENTION: CYTOCHROME P450 MONOOXYGENASE AND NADPH CYTOCHROME  
; TITLE OF INVENTION: P450 OXIDOREDUCTASE  
; TITLE OF INVENTION: GENES AND PROTEINS RELATED TO THE OMEGA HYDROXYLASE COMPLEX OF C  
; TITLE OF INVENTION: TROPICALIS AND METHODS RELATING THERETO  
; FILE REFERENCE: 1010-16  
; CURRENT APPLICATION NUMBER: US/10/139,218  
; PRIOR FILING DATE: 2002-05-03  
; PRIOR FILING DATE: US/09/976,800  
; NUMBER OF SEQ ID NOS: 118  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 100  
; LENGTH: 517  
; TYPE: PRT  
; ORGANISM: CANDIDATROPICALIS  
US-10-139-218-100

Query Match 12.9%; Score 396; DB 14; Length 517;  
Best Local Similarity 26.6%; Pred. No. 3.5e-32;  
Matches 139; Conservative 78; Mismatches 162; Indels 144; Gaps 19;  
QY 93 TYIA-----DPNVVHVLKTNFTNPKGIVYRSYMDVLLGDGIFNADGELWRKQ 141  
Db 84 TVVSILFGTRIVVTKDPENIKAILATQFGDFSLGKRHTLFKP--LLGDGIFTLDDGEGWKHS 142  
QY 142 RKTASFEFASKNLRDFAI-----VFREYSILKSLGILSOASKAGKVVDMQELYMRMTLDS 196  
Db 143 RAMLRPQFAREQVAHVTSLEPHQLKKHLK-----HKGEYFDIQELFFRFTVDS 193  
QY 197 ICKVGVGEIGTSLPD-----LPENSFAQAFDAANIITLRFIDPLWRIKRFPH 245

Db 194 ATEFLGESVHSLKDESIGINQDDIDFAGRKDFAESFNKAQEYLAIRTL-----VQTFYV 248  
QY 246 VGSEALLAQSILKLVDEFT--YSVIRRRKAEIIVEV--RASG-----KQKMKHDLISRFI 295  
Db 249 LVNNKEFRDCTKLHKFTNYVYQKALDASPEELEKQSGVVFYELVKQTR-----298  
QY 296 ELGEAGDDGGGDDKSLRDVNLVNFVIAGRDTTATTLTSMFTMAMSHPDVAEKLRELCA 355  
Db 299 -----DENVLRDQSLNILLAGRTDTTAGLLSFAVFLARHPETWAKLREI--343  
QY 356 FEARAREEGVTILVLCGGADADKAFARVAQAGLLTVDSLGKLVHLHACVTETLRLYP 415  
Db 344 -----EQQFGL-----GED-----SRVEE-----ITFSLKRCBYLKAFLNETLRIYP 381  
QY 416 AVPDQPKGILEDDVLP-----DGTK-----VRAGMVTYVPVYSGMGEYNNWGPDAASFRPER 467  
Db 382 SVERNFRIATKNTLPRGGSGDGTSPILQKGAVSYGINSHTLDPVYVGPDAAEFRPER 441  
QY 468 WINEDGAFRNASPFKFTAFQAGPRICLGKDSAYLQMKMALAILFRFYSFLLLEGHPVQYR 527  
Db 442 WPEPS---TKLGLWYLPFNGGPRICLGQOFA-----470  
QY 528 MMTILSMAHGLKVRVSRAVCHGDLMDIV--PLNPRQITILVLQ 568  
Db 471 ----LTEAGYVLRVLVQEFSHVLDPEVYPPKRLTNLTWCLQ 509

RESULT 14  
US-10-400-902-32  
; Sequence 32, Application US/10400902  
; Publication No. US20030175896A1  
; GENERAL INFORMATION:  
; APPLICANT: Craft, David L.  
; APPLICANT: Wilson, C. Ron  
; APPLICANT: Erich, Dudley  
; APPLICANT: Zhang, Yeyan  
; TITLE OF INVENTION: USE OF CYP52A2A PROMOTER TO INCREASE GENE EXPRESSION IN YEAST  
; FILE REFERENCE: U0012 OS/OAAP (1010-49)  
; CURRENT APPLICATION NUMBER: US/10/400,902  
; CURRENT FILING DATE: 2003-03-31  
; PRIOR APPLICATION NUMBER: US/09/911,781  
; PRIOR FILING DATE: 2001-07-24  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 32  
; LENGTH: 517  
; TYPE: PRT  
; ORGANISM: Candida tropicalis  
US-10-400-902-32

Query Match 12.9%; Score 396; DB 14; Length 517;  
Best Local Similarity 26.6%; Pred. No. 3.5e-32;  
Matches 139; Conservative 78; Mismatches 162; Indels 144; Gaps 19;  
QY 93 TYIA-----DPNVVHVLKTNFTNPKGIVYRSYMDVLLGDGIFNADGELWRKQ 141  
Db 84 TVVSILFGTRIVVTKDPENIKAILATQFGDFSLGKRHTLFKP--LLGDGIFTLDDGEGWKHS 142  
QY 142 RKTASFEFASKNLRDFAI-----VFREYSILKSLGILSOASKAGKVVDMQELYMRMTLDS 196  
Db 143 RAMLRPQFAREQVAHVTSLEPHQLKKHLK-----HKGEYFDIQELFFRFTVDS 193  
QY 197 ICKVGVGEIGTSLPD-----LPENSFAQAFDAANIITLRFIDPLWRIKRFPH 245  
Db 194 ATEFLGESVHSLKDESIGINQDDIDFAGRKDFAESFNKAQEYLAIRTL-----VQTFYV 248  
QY 246 VGSEALLAQSILKLVDEFT--YSVIRRRKAEIIVEV--RASG-----KQKMKHDLISRFI 295  
Db 249 LVNNKEFRDCTKLHKFTNYVYQKALDASPEELEKQSGVVFYELVKQTR-----298  
QY 296 ELGEAGDDGGGDDKSLRDVNLVNFVIAGRDTTATTLTSMFTMAMSHPDVAEKLRELCA 355  
Db 299 -----DENVLRDQSLNILLAGRTDTTAGLLSFAVFLARHPETWAKLREI--343

QY 356 FEAEARAREEGVTLLVLCGGADADDKAFARVAQFAGLLTYDSLGLVYLHACVTETRLYP 415  
Db 344 -----EQQFGL-----SRVEE-----ITFESLKRCEYLKAFINETLRIYP 381  
QY 416 AVPQDPKGILLEDDVLP-----DGTK---VRAGGMVTVVPYSMGRMEYNWGPDAAFRPER 467  
Db 382 SVPRNFRIATKNTTLPGRGGSDGTSPILIQKEAVSYGINSTHLDPPVYGGDAAAEFRPER 441  
QY 468 WINEDGAFRNASPFKFTAFQAGPRICLQKDSAYLQMKWALAILFRFYSFRLLLEGHPVQYR 527  
Db 442 WFEPS---TKKLWAYLPFNGGPRICLQQA----- 470  
QY 528 MMTILSMAGHLKVRVSRVAVCHGDLMDIV--PLNPRQITVLVQ 568  
Db 471 ----LTEAGYVLRLVQEFHSHVRLDPDEVYPPKRLTNLTMCLO 509

Search completed: March 2, 2004, 15:34:53  
Job time : 40 secs

QY 356 FEAEARAREEGVTLLVLCGGADADDKAFARVAQFAGLLTYDSLGLVYLHACVTETRLYP 415  
Db 344 -----EQQFGL-----SRVEE-----ITFESLKRCEYLKAFINETLRIYP 381  
QY 416 AVPQDPKGILLEDDVLP-----DGTK---VRAGGMVTVVPYSMGRMEYNWGPDAAFRPER 467  
Db 382 SVPRNFRIATKNTTLPGRGGSDGTSPILIQKEAVSYGINSTHLDPPVYGGDAAAEFRPER 441  
QY 468 WINEDGAFRNASPFKFTAFQAGPRICLQKDSAYLQMKWALAILFRFYSFRLLLEGHPVQYR 527  
Db 442 WFEPS---TKKLWAYLPFNGGPRICLQQA----- 470  
QY 528 MMTILSMAGHLKVRVSRVAVCHGDLMDIV--PLNPRQITVLVQ 568  
Db 471 ----LTEAGYVLRLVQEFHSHVRLDPDEVYPPKRLTNLTMCLO 509

RESULT 15  
US-10-405-660-100  
; Sequence 100, Application US/10405660  
; Publication No. US20030186411A1  
; GENERAL INFORMATION:  
; APPLICANT: Wilson, Ron C.  
; APPLICANT: Craft, David L.  
; APPLICANT: Eirich, Dudley  
; APPLICANT: Eshoo, Mark  
; APPLICANT: Madduri, Krishna M.  
; APPLICANT: Cornett, Cathy A.  
; APPLICANT: Brenner, Alfred A.  
; APPLICANT: Tang, Maria  
; APPLICANT: Loper, John C.  
; APPLICANT: Gleeson, Martin  
; TITLE OF INVENTION: CYTOCHROME P450 MONOOXYGENASE AND NADPH CYTOCHROME P450 OXIDOREDUCTASE  
; TITLE OF INVENTION: GENES AND PROTEINS RELATED TO THE OMEGA HYDROXYLASE COMPLEX OF CYTOCHROME P450  
; TITLE OF INVENTION: TROPICALIS AND METHODS RELATING THERETO  
; FILE REFERENCE: 1010-16  
; CURRENT APPLICATION NUMBER: US/10/405,660  
; CURRENT FILING DATE: 2003-04-02  
; PRIOR APPLICATION NUMBER: US/09/976,800  
; PRIOR FILING DATE: 2001-10-12  
; NUMBER OF SEQ ID NOS: 118  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 100  
; LENGTH: 517  
; TYPE: PRT  
; ORGANISM: CANDIDATROPICALIS  
US-10-405-660-100

Query Match 12.9%; Score 396; DB 14; Length 517;  
Best Local Similarity 26.6%; Pred. No. 3.Se-32;  
Matches 139; Conservative 78; Mismatches 162; Indels 144; Gaps 19;

QY 93 TYIA-----DPNVHVLKNTFTNPKGIVYRSMVLLGDIENADGELWRKQ 141  
Db 84 TYVILFGRIVTVKDPENIKAILATQGFSLGKRTLFXP-LLGDGIFTLDEGKWS 142  
QY 142 RKTASFEFFASKNLRDSAL-----VPREYSLKLSGILSQASKAGKVVDMQELYNMTLDS 196  
Db 143 RAMLRPOQFAREQVAHVTSLEPHFQLKKHLK-----HKGEYFDIQELFFRFTVDS 193  
QY 197 ICKVGFGEIGTLSPD-----LPNSFAQADAAIIITLRFIDPLWRIKEFEH 245  
Db 194 ATEFLFGESVHSLKDESGINQDDIDFAGRKDXFAESFNKAQEYLAIRTL-----VQTFYW 248  
QY 246 VGSEALLAQSTKLYDEFT-YSVIRRRKAEIIEV-RASG-----KQEKMKHDIILSRFI 295  
Db 249 LVNNKFRDCTKLHFKFTNYTVQKALDASPEELEKQSGYVFLYELVKQTR----- 298  
QY 296 ELGEAGDGGGFGDDKSLRVNLFVIAGRDRTATTLSWFTMAMSHPDVAEKLRLCA 355  
Db 299 -----DPNVLRDQSLNILLAGRTTAGLLSFVAFELARHPEIWKLRREEI-- 343





Qy	356	FEABRAREGVTVL	CGGADDDKAFARVAQFAGLLTYDSIGKLVYHACVTELTIRYP	415
		:-	:-	
Db	344	-----EQQFGL	-----GED-----SRVEE-----ITFSELKRCYTLKAFLNELTIRYP	381
		:-	:-	
Qy	416	AVPQDPKGILEDDVLP	-----DGTG---VRAGMYTVYPYSGRMRYNKGWPDAAASFRPER	467
		:-	:-	
Db	382	SVPRNFRATKNTL	PRGGSDGTSPILIQGEAVSYGINSTHLPDVPYVYGGPDAAASFRPER	441
		:-	:-	
Qy	468	WINEDGAFARNASP	KFTAFQAGPRICLGDKDSAYLQMKVALILFRFYSPRLLEGHGPVQR	527
		:-	:-	
Db	442	WFEESS---TKKLGWAYLP	ENGPGPRICLGGQFA-----	470
		:-	:-	
Qy	528	NMTILSMAHGKLV	RSRVAVCHGDLMDIV-PLNPRQTLVLQ	568
		:-	:-	
Db	471	---LTEAGVYLV	IVQFESSHVLRLDPDEVYPPKRLTNLTMCLO	509
		:-	:-	

## RESULT 4

```

US-09-911-781-32
RESOLUTION 4
; Sequence 32, Application US/09911781
; Patent No. 6673613
; GENERAL INFORMATION:
; APPLICANT: Craft, David L.
; APPLICANT: Wilson, C. Ron
; APPLICANT: Ehrich, Dudley
; APPLICANT: Zhang, Yeyan
; TITLE OF INVENTION: USE OF CYP5A2A PROMOTER TO INCREASE GENE EXPRESSION IN YEAST
; FILE REFERENCE: U0012 OS/OAAP (1010-49)
; CURRENT APPLICATION NUMBER: US/09/911,781
; CURRENT FILING DATE: 2001-07-24
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 32
; LENGTH: 517
; TYPE: PRT
; ORGANISM: Candida tropicalis
US-09-911-781-32

```

[illegible]

```

QY 528 MWTLISMAHGLKVRVSRVAVCHGDLMDIV--PLNPRQITLVLQ 568
      : : : : : : : : : : : : : : : : : : : :
pb 471 ----LTEAGYVLVRLVOEFSHVRLDPEVYPPKRLTLTMCLO 509

```

## RESULT 5

```

US-09-302-620B-101
; Sequence 101, Application US/09302620B
; Patent No. 6331420
; GENERAL INFORMATION:
; APPLICANT: Wilson, C. Ron
; APPLICANT: Craft, David L.
; APPLICANT: Ehrich, Dudley
; APPLICANT: Eshoo, Mark
; APPLICANT: Madduri, Krishna M.
; APPLICANT: Cornett, Cathy A.
; APPLICANT: Brenner, Alfred A.
; APPLICANT: Tang, Maria
; APPLICANT: Loper, John C.
; APPLICANT: Gleeson, Martin
; TITLE OF INVENTION: CYTOCHROME P450 MONOOXYGENASE AND NADPH CYTOCHROME P450
; TITLE OF INVENTION: OXIDOREDUCTASE GENES AND PROTEINS RELATED TO THE OMEGA
; TITLE OF INVENTION: HYDROXYLASE COMPLEX OF CANDIDA TROPICALIS AND METHODS
; TITLE OF INVENTION: RELATING THERETO
; FILE REFERENCE: 1010-16 seq
; CURRENT APPLICATION NUMBER: US/09/302,620B
; CURRENT FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 109
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 101
; LENGTH: 517
; TYPE: PRT
; ORGANISM: Candida tropicalis
US-09-302-620B-101

```

Query Match	12.8%;	Score 395;	DB 4;	Length 517;	
Best Local Similarity	27.5%;	Pred. No. 6.1e-35;			
Matches 132;	Conservative 77;	Mismatches 171;	Indels 100;	Gaps 16;	
QY	90	TSYTYIADPNVNEHLKTNFNPNYPKGIYRSYMDVLGGDGIFFNADGELWRKQRTASPEF	149		
DB	92	TKIVVTKDPENIKAILATQFGDPSLGGKHALFKP-LJGEGIFTLDGEGKHRSMLRQF	150		
QY	150	ASKNLRDPSAI-----VREYSKLKSLGILSOASKAGKGVVDMQELVYMRMTLDSICKVGRGV	204		
DB	151	AREQVAHVTSLEPHQLKXHLK-----HKGEYFDIQELFFRFTVDSAGTEFLGE	201		
QY	205	EIGTLSPD-----LPENSGFAQAFDAANIITLFFIDPLWRIKRRFFHVSEALLA	253		
DB	202	SVHSLKDEITICINQDDIDFAGEKQFAESFNKAQEYLSIFIL-----VQTEYWLINKEFR	256		
QY	254	QSIKLVDEFT-YSVTRRRKAIIVEVRASG-----KQMKKHDIILSRLETGEADD	303		
DB	257	DCTKLIVHKEFTNYQKALDAPTEELKQGGVYFELYELVKQTR-----	298		
QY	304	GGGFGDDKSLRDVNLVFIAGRDITATLTSWFETHMANGHPDVAEKLRELCAFEAREARE	363		
DB	299	-----DPKVLRDQSLNILLAGDITTAGLLSPAVFELARNPHIWLKREEI-----E	344		
QY	364	EGVTVLVCGGADADKFAARVAQPAAGLLTVDSLGKLYVLHACVTETIELPVPQDPKG	423		
DB	345	QQQFGL-----GED-----SRVEE-----ITFESLKRCEYLKAFNLNETLURVPSVPERNFRI	389		
QY	424	ILEDDVL-----PDGTK-----VRAGGMVTVYPYSMGRMEYNWGPDAASFRPRWRINEDGAF	475		
DB	390	ATKNLTLPRGGPDGTQITLQKGEVSYGINSIHLDPVYYPGPDAAEFRPRWPFPS---	446		
QY	476	RNASPFKFTAFQAGPRICLGKDSAYLQMKMALILFRFYS-PRLLEGHPVQVRMTILSM	534		
DB	447	TRKLGWALYFPNGGPRICLGGQFALTEAGYVILRVLQVBFSHIRLDPDEYVPPKRLTNLTM	506		

Sequence 95, Application US/09302620B  
Patent No. 6331420  
GENERAL INFORMATION:  
APPLICANT: Wilson, C. Ron  
APPLICANT: Craft, David L.  
APPLICANT: Eirich, Dudley  
APPLICANT: Eshoo, Mark  
APPLICANT: Madduri, Krishna M.  
APPLICANT: Cornett, Cathy A.  
APPLICANT: Brenner, Alfred A.  
APPLICANT: Tang, Maria  
APPLICANT: Loper, John C.  
APPLICANT: Gleeson, Martin  
TITLE OF INVENTION: CYTOCHROME P450 MONOOXYGENASE AND NADPH CYTOCHROME P450  
TITLE OF INVENTION: OXIDOREDUCTASE GENES AND PROTEINS RELATED TO THE OMEGA  
TITLE OF INVENTION: HYDROXYLASE COMPLEX OF CANDIDA TROPICALIS AND METHODS  
TITLE OF INVENTION: RELATING THERETO  
FILE REFERENCE: 1010-16.seq  
CURRENT APPLICATION NUMBER: US/09/302,620B  
CURRENT FILING DATE: 1999-04-30  
NUMBER OF SEQ ID NOS: 109  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 95  
LENGTH: 523  
TYPE: PRT  
ORGANISM: Candida tropicalis  
US-09-302-620B-95

Query Match 12.6%; Score 389; DB 4; Length 523;  
Best Local Similarity 28.1%; Pred. No. 2.9e-34;  
Matches 127; Conservative 60; Mismatches 167; Indels 98; Gaps 16;

QY 97 DPVNVHVLKNTFTYKIVYRSYMDVLGDFINAGELWKRKQKTAFSEFASKNLRD 156  
DB 105 DENIKAVLATQTFDSLG-TRHAFAPLLGIGITLDGEGWKHGRAMLRPOFARDQIGH 163  
QY 157 PSA---IVFREYSKLKSGILSKQASKAGKQVDMQELMYRMITLDSICKVGFVGEIGTL--- 209  
DB 164 VKALPEPHQIMAKQIKLN-----QKTFDIQELFFRTVDATTEFLFGESVHSLYDE 215  
QY 210 -----SP-DLP-ENSPFAQAFDAANTIIITLRFIDP-----LWRIKRFPHVGESEALLAQSIKL 258  
DB 216 KLGIPTPEIPGRENFAAFNVQHYLATRSYSQTFYFLTNPKFPRDCN-----AKVHHL 270  
QY 259 VDEFTYSVIRRKALIVEVRASG-----KQKMKHDLISRFTELGEAGDDGGFGDD 310  
DB 271 AKYFVNKALNFTPEELEKSKSGYVFLYELVKQTR-----DP 307  
QY 311 KSLRDVNLNFIAGDRTTATTLNFWTHMAMSHPDVAEKLRLRELCAFEARAREEGVTLLVL 370  
DB 308 KVLQDOLLNIMVAGDRTTAGLLSFALFELARHPKMWKSLREI-----EVNFGV----- 356  
QY 371 CGGADADDKAFARVAQFAGLTYSLSGLKLVHACVTEITLILYPAVPODPKGILEDVIL 430  
DB 357 --GED-----SRVBE-----ITFEALKRCEYLKALNETLMYSPVNFRTATRDITL 403  
QY 431 P-----DGKTVRAGGMVTVYPSGMGMEYNWGDAAASFRPERMINEDGAFRNASPFK 482  
DB 404 PRGGGANGTDPIYIPKSGTVAVVVYKTHRELYEYKGDANDFRPERWFPS-----TKKLGWA 460  
QY 483 FTAFQAGPRICLKDSAYLQMKMALAILFRFY 514  
DB 461 YVFPNGPRVCLGQFALTEASVITRLAQMF 492

RESULT 8  
US-09-302-620B-96  
Sequence 96, Application US/09302620B  
Patent No. 6331420  
GENERAL INFORMATION:  
APPLICANT: Wilson, C. Ron  
APPLICANT: Craft, David L.  
APPLICANT: Eirich, Dudley

Sequence 97, Application US/09302620B  
Patent No. 6331420  
GENERAL INFORMATION:  
APPLICANT: Wilson, C. Ron  
APPLICANT: Craft, David L.  
APPLICANT: Eirich, Dudley  
APPLICANT: Eshoo, Mark  
APPLICANT: Madduri, Krishna M.  
APPLICANT: Cornett, Cathy A.  
APPLICANT: Brenner, Alfred A.  
APPLICANT: Tang, Maria  
APPLICANT: Loper, John C.  
APPLICANT: Gleeson, Martin  
TITLE OF INVENTION: CYTOCHROME P450 MONOOXYGENASE AND NADPH CYTOCHROME P450  
TITLE OF INVENTION: OXIDOREDUCTASE GENES AND PROTEINS RELATED TO THE OMEGA  
TITLE OF INVENTION: HYDROXYLASE COMPLEX OF CANDIDA TROPICALIS AND METHODS  
TITLE OF INVENTION: RELATING THERETO  
FILE REFERENCE: 1010-16.seq  
CURRENT APPLICATION NUMBER: US/09/302,620B  
CURRENT FILING DATE: 1999-04-30  
NUMBER OF SEQ ID NOS: 109  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 97  
LENGTH: 522  
TYPE: PRT  
ORGANISM: Candida tropicalis  
US-09-302-620B-97

Query Match 12.8%; Score 393; DB 4; Length 522;  
Best Local Similarity 28.4%; Pred. No. 1e-34;  
Matches 140; Conservative 73; Mismatches 184; Indels 96; Gaps 20;

QY 77 LSHRVTVTVDMPSTYIIA--DPNVHVLKNTFTYKIVYRSYMDVLGDFINAD 134  
DB 82 LNPDPITPTFFPESINLSTLEPENIKALLATQFNDPSLG-TRHSHFAPLLGIGITLD 140  
QY 135 GELWKRQKTAFSEFASKNLRDPSAIVFREYSKLKSGILSKQASKA-GKVVDMQELMYRMT 193  
DB 141 GAGWKHSRMLRQFAREQI---SHVKLLPEPHQV---FFKHVRKAQKQTFDIQELFFRLT 195  
QY 194 LDSICKVGGVEIGTSPDLPENSPFAQAFDAANTIIITLRFIDPILWIKRFFHVGSEALLA 253  
DB 196 VDSATBFLFGESVESLR---DESIGMSINALDFDGKAGFADAFNYSQNY--LASAVMQ 249  
QY 254 QSIKLVDEFTYSVIRRKALIVEVRASGKQKMKHDLISRFIELGE---AGDDGGGF--- 307  
DB 250 Q-----IYWLNGKKFK-----ECNAKVHKADYVVS KALLDTPEQLEKQDGYVFLYE 297  
QY 308 -----GDKSLRDVNLNFIAGDRTTATTLNFWTHMAMSHPDVAEKLRLRELCAFEARAR 362  
DB 298 LVKQTRDRQVLRDQLNIMVAGDRTTAGLLSFVFFELARNPEVTNKLREI----- 348  
QY 363 BEGTVILCGADADDK---AFAARVAQFAGLTYSLSGLKLVHACVTEITLILYPAVPO 419  
DB 349 -----EDKFGLENARVED-----ISFESLKCEYLKAVINETLILYSPVQ 390  
QY 420 DPKGILEDDVLDPG-----TKVRAGGMVTVYPSGMGMEYNWGDAAASFRPERWINE 471  
DB 391 NFRVATKNTTLPRGGGKDGSLPVLVRKQTVMYGVYAAHNPVYVYKDALEFRPERWFEP 450  
QY 472 DGAFNASPFKFTAFQAGPRICLKDSAYLQMKMALAILFRFYFRILLE--GH-----PV 524  
DB 451 E---TKKLGWAFPLFNGGPRICLKQFALTEAS-----YVTVRLIQBFHGLSMDPNT 499  
QY 525 QY--RMTILSMA 535  
DB 500 EYPPRKMSHLTMS 512

RESULT 7  
US-09-302-620B-95

Mon Mar 8 10:13:27 2004

```

; APPLICANT: Eshoo, Mark
; APPLICANT: Madduri, Krishna M.
; APPLICANT: Cornett, Cathy A.
; APPLICANT: Brenner, Alfred A.
; APPLICANT: Tang, Maria
; APPLICANT: Loper, John C.
; APPLICANT: Gleeson, Martin
; TITLE OF INVENTION: CYTOCHROME P450 MONOOXYGENASE AND NADPH CYTOCHROME P450
; TITLE OF INVENTION: OXIDOREDUCTASE GENES AND PROTEINS RELATED TO THE OMEGA
; TITLE OF INVENTION: HYDROXYLASE COMPLEX OF CANDIDA TROPICALIS AND METHODS
; TITLE OF INVENTION: RELATING THERETO
; FILE REFERENCE: 1010-16.seq
; CURRENT APPLICATION NUMBER: US/09/302,620B
; CURRENT FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 109
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 96
; LENGTH: 522
; TYPE: PRT
; ORGANISM: Candida tropicalis
US-09-302-620B-96

Query Match      12.6%; Score 387; DB 4; Length 522;
Best Local Similarity 27.3%; Pred. No. 4.9e-34;
Matches 131; Conservative 69; Mismatches 185; Indels 94; Gaps 16;

QY 68 RMHDLVGLSRHRTVTVDMPFTSYTYI--ADPVNVEHVLKTNFTNYPKGIYVRSYMDVL 125
Db 78 RIHD-----LDRPDIPFTFPVFSINLVNTLEPENIKAILATQFNDFSLG-TRHSHFAPL 131
QY 126 LGDGFNADGELMRKQRTKATSFASAKNLRDPSAI-----VFREYSKLGLSQASKAG 180
Db 132 LGDGFITLDGAGWKHSRMLRPOFARQIASHVKLLPHVQVFFKHVKAQ-----G 182
QY 181 KVTDMQELVMTLDSICKVGFVEIGTLPDLPENSFAQAFDAANIITLRFIDPLMRI 240
Db 183 KTFDIQELFRLTVDSTATEFLFGESVSLR-----DESIGSINALDGDGAGFADAFNYS 238
QY 241 KRFFHVSEALLAQSISKLVDEFTYSVIRRRKAEIVVRASQKQKMHDLRSFIELGE- 299
Db 239 QNY--LASRAVWQ-----LYWVLNGKFKF---ECNAKVKFADYYVYNKALDLTPE 284
QY 300 --AGDGGGF-----GDDKSRDVLNVIAGRTTATLTSWFTMHMNSHPDVAEKL 349
Db 285 QLEKQDGVFLYELVKQTRDKQVLRDQLLNTMVAGRDRTTAGLLSFVFFELARNPEVTNKL 344
QY 350 RRELCAFEARAEEGVTNLVLCGGADADDKAFARVAQFAGLLTYDSLGKLVYLHACVTE 409
Db 345 REEI-----EDKFLGENASVED-IGFESLSKCEYLKAVLINE 380
QY 410 TRLYPVPQDPKGILEDVLPDG-----TKVRAGGVYTYVPYSGRMVYKNGPDAA 461
Db 381 TRLYPSPQPNFRVATKNTLTPRGGKDGSLGPVLVRKQTVIYGVYAAHRNPVAVYKDAL 440
QY 462 SFRERWINEGAFRNASPFKFTAFQAGPRICLQKDSAYLQMKWALALFLRFYSFRLLE 520
Db 441 EFRPERWFEPE---TKLWGAFLPENGPGPRICLQGOQFALTEAS-----YTVRLLIQ 488

RESULT 9
US-09-302-620B-98
; Sequence 98, Application US/09302620B
; Patent No. 6331420
; GENERAL INFORMATION:
; APPLICANT: Wilson, C. Ron
; APPLICANT: Craft, David L.
; APPLICANT: Eirich, Dudley
; APPLICANT: Eshoo, Mark
; APPLICANT: Madduri, Krishna M.
; APPLICANT: Cornett, Cathy A.
; APPLICANT: Brenner, Alfred A.
; APPLICANT: Tang, Maria
; APPLICANT: Loper, John C.

```

```

; APPLICANT: Gleeson, Martin
; TITLE OF INVENTION: CYTOCHROME P450 MONOOXYGENASE AND NADPH CYTOCHROME P450
; TITLE OF INVENTION: OXIDOREDUCTASE GENES AND PROTEINS RELATED TO THE OMEGA
; TITLE OF INVENTION: HYDROXYLASE COMPLEX OF CANDIDA TROPICALIS AND METHODS
; TITLE OF INVENTION: RELATING THERETO
; FILE REFERENCE: 1010-16.seq
; CURRENT APPLICATION NUMBER: US/09/302,620B
; CURRENT FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 109
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 98
; LENGTH: 540
; TYPE: PRT
; ORGANISM: Candida tropicalis
US-09-302-620B-98

Query Match      12.5%; Score 383.5; DB 4; Length 540;
Best Local Similarity 25.5%; Pred. No. 1.3e-33;
Matches 148; Conservative 87; Mismatches 209; Indels 137; Gaps 23;

QY 5 REAHLTPA---TPSPFPPLAGBHKYIALLLVLSWLVQVWSLRKQK---PRSWPVIKA 58
Db 53 RRFHAKPLGNFVRDPTFGIATP-----LLLIYL-----KSGITVMKFAWGL--- 93
QY 59 TVBQLRNVRHMDWLVGYLSR---HRTVTVDMEFTSYTYIADPVNVEHVLKTNFTNYPKG 115
Db 94 -----WNKVIYVDPKYYKTGLRIVGLPLIETMDPENIKAVLATQFNDFSLG 140
QY 116 IVRSYMDVLGDIIFNADGELMRKQRTKATSFASAKNLRDPSAI-----VFREYSKLKS 170
Db 141 -TRHDFLYSLGDIIFTLDGAGWKHSRMLRPOFARQVSHVKLLPHVQVFFKHVKA-- 197
QY 171 GILSQASKAKGVDMQELVMTLDSICKVGF-----VEIGTLPDLP-----NSF 218
Db 198 -----HRGQTDFIQELFRLTVDSTATEFLFGESASLSRDESIG-LTPTKDFDGRDF 249
QY 219 AQAPDAANIITLRFIDPLMRIKRFFHVSEALLAQSISKLVDEFTYSVIRRRKAEIVEVR 278
Db 250 ADAFNYSQTYQAYRFL-----LQMYWILNGSBERKSIADVHKFADHYV-QKALELTD-- 301
QY 279 ASGQKQKMHDLISRFIELGEAGDGGGFDGDKSLRDVLNFIAGBDTATTLSWFTMH 338
Db 302 ---DDLQKQGVVFLYELAKQ-----TRDPKVLRLQLLNIIIVAGRDTTAGLLSFVFE 351
QY 339 AMGHPDVAEKLRELCAFEARAEERAREEGVTNLVLCGGADADDKAF---ARVAQFAGLLTY 394
Db 352 LSRNPEVFAKLREBVE-----ENRFGLGEARVEE-----ISF 383
QY 395 DSLGKLVYLHACVTEITLRLYPVPQD-----PKGILEDDVLPDGTKVRAGGVMTY 444
Db 384 ESLKSCYELKAVINETLRLYPSVPHNFRVATRNITLPRGGGEDGYSP--IWKKGQVVMY 441
QY 445 VPSYSGRMVYKNGPDAAFRPERWINEGAFRNASPFKFTAFQAGPRICLQKDSAYLQMK 504
Db 442 TVIATHRDPISYIGADADVFRPERWFEPE---TRKLWYVYFPNGGPRICLQGOQFALTEAS 498
QY 505 MALAILFRFYSFRLLE---GHPVQYRMVMTLSMAHGLKVRV 542
Db 499 YTVRLLIQEFAHUSMDPDTTEYPPKLQNTLTLSLFDGADVNM 539

RESULT 10
US-09-302-620B-102
; Sequence 102, Application US/09302620B
; Patent No. 6331420
; GENERAL INFORMATION:
; APPLICANT: Wilson, C. Ron
; APPLICANT: Craft, David L.
; APPLICANT: Eirich, Dudley
; APPLICANT: Eshoo, Mark
; APPLICANT: Madduri, Krishna M.
; APPLICANT: Cornett, Cathy A.
; APPLICANT: Brenner, Alfred A.

```

```

; APPLICANT: Tang, Maria
; APPLICANT: Brenner, John C.
; APPLICANT: Gleeson, Martin
; TITLE OF INVENTION: CYTOCHROME P450 MONOOXYGENASE AND NADPH CYTOCHROME P450
; TITLE OF INVENTION: OXIDOREDUCTASE GENES AND PROTEINS RELATED TO THE OMEGA
; TITLE OF INVENTION: HYDROXYLASE COMPLEX OF CANDIDA TROPICALIS AND METHODS
; TITLE OF INVENTION: RELATING THERETO
; FILE REFERENCE: 1010-16.seq
; CURRENT APPLICATION NUMBER: US/09/302,620B
; CURRENT FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 109
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 102
; LENGTH: 512
; TYPE: PRT
; ORGANISM: Candida tropicalis
US-09-302-620B-102

```

```

Query Match 12.4%; Score 381; DB 4; Length 512;
Best Local Similarity 26.1%; Pred. No. 2.2e-33;
Matches 145; Conservative 89; Mismatches 181; Indels 140; Gaps 24;

```

```

QY 26 YIAL-LLVLSWIL--VORSLRKQKGRSPVPI-----GATVEQLRNYHRMDW 72
DB 10 YIVLELLAIQIVAHVTRNYLMKLGAKPFTHVQDGMWLGKFGREFLKAAGSLVDL 69
QY 73 LVGLSRHRTVDMPTFSY-----TYIADPNVNEHVILKTNFTNPKGIVYRSTMDVLL 126
DB 70 IISRFHEDNT-----FSSYAFGNHVFTROPENIKALLATQFGDFSLGSRVK-FFKPLL 123
QY 127 GDGIFNADGELWRKQKRTASFFASKNLRDPSAI-----VREYSLKLSGLISQASKAGK 181
DB 124 GYGIFTLDAEGWKHSRAMLRPFQAREQVAHVTSLEPHFQLKKHLK-----HKGE 174
QY 182 VDMQELMYRMTLDSICKVGF-----VEIGTILSPDLPE--NSFAQAFDAANIITL 231
DB 175 YFDIQELFFRTVDSATEFLFESVHSLRDEEIGYDTKDMAEERRKPADAFNKSQVYVAT 234
QY 232 RFIDPLMRKFRFFHVGSEALLAQSIKLVDFT--YSVIRRRKAEIIVEVRASG-----281
DB 235 RV-----ALONTYLVNKNKEFCNDIVHKFTNYVQKALDTPBELEKQGGVFLYELV 289
QY 282 KQEKMKHILSRFIELGEAGDGGGDKSLRDVVLNFIAGRDTTATTLTTSWFTHMAMS 341
DB 290 KQTR-----DPKVLDDQSLNILLAGRDTTAGLLSFAVFEELAR 326
QY 342 HPDVAEKLRLRELCAPAEAREBEGVTLVLCGGADADDKAFARVAQFAGLLTYDSLGLV 401
DB 327 NPHIWAQLREEI-----EQQFGL-----GED-----SRVEE-----ITFESLKRCE 362
QY 402 YLHACVTTETRLYPVAVPQDPKGILEDDVLPDGTQKVRAGG-----MVTYVPYSMGR 451
DB 363 YLKAVLNELTLRHPSVPRNARPAIKDTTLP-----RGGPGNGKDPILIRKDEVVQYSISA 417
QY 452 MEYN---WGPDAASFRPERWINEGDAFRNASPFKFTAFQAGPRICLGKDSAYLQMKWALA 508
DB 418 TQTNPAVYAGDAADFRPERWF--EPSTRNLG-WAFLPFGNGGPRICLGQGFALTEAGYVLV 474
QY 509 ILFRFYSFRLLEGHP 523
DB 475 -----RLVQEF 481

```

```

RESULT 11
US-09-302-620B-103
; Sequence 103, Application US/09302620B
; Patent No. 6331420
; GENERAL INFORMATION:
; APPLICANT: Willson, C. Ron
; APPLICANT: Craft, David L.
; APPLICANT: Birch, Dudley
; APPLICANT: Eschoo, Mark
; APPLICANT: Magduri, Krishna M.

```

```

; APPLICANT: Cornett, Cathy A.
; APPLICANT: Brenner, Alfred A.
; APPLICANT: Tang, Maria
; APPLICANT: Loper, John C.
; APPLICANT: Gleeson, Martin
; TITLE OF INVENTION: CYTOCHROME P450 MONOOXYGENASE AND NADPH CYTOCHROME P450
; TITLE OF INVENTION: OXIDOREDUCTASE GENES AND PROTEINS RELATED TO THE OMEGA
; TITLE OF INVENTION: HYDROXYLASE COMPLEX OF CANDIDA TROPICALIS AND METHODS
; TITLE OF INVENTION: RELATING THERETO
; FILE REFERENCE: 1010-16.seq
; CURRENT APPLICATION NUMBER: US/09/302,620B
; CURRENT FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 109
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 103
; LENGTH: 512
; TYPE: PRT
; ORGANISM: Candida tropicalis
US-09-302-620B-103

```

```

Query Match 12.3%; Score 380; DB 4; Length 512;
Best Local Similarity 26.1%; Pred. No. 2.9e-33;
Matches 145; Conservative 87; Mismatches 183; Indels 140; Gaps 24;

```

```

QY 26 YIAL-LLVLSWILVQRWS--LRKQKGRSPVPI-----GATVEQLRNYHRMDW 72
DB 10 YIVLPLLVITKQIVAHARTNYLMKLGAKPFTHVQDGMWLGKFGREFLKAAGSLVDL 69
QY 73 LVGLSRHRTVDMPTFSY-----TYIADPNVNEHVILKTNFTNPKGIVYRSTMDVLL 126
DB 70 IISRFHEDNT-----FSSYAFGNHVFTROPENIKALLATQFGDFSLGSRVK-FFKPLL 123
QY 127 GDGIFNADGELWRKQKRTASFFASKNLRDPSAI-----VREYSLKLSGLISQASKAGK 181
DB 124 GYGIFTLDDGEMKHSRAMLRPFQAREQVAHVTSLEPHFQLKKHLK-----HKGE 174
QY 182 VDMQELMYRMTLDSICKVGF-----VEIGTILSPDLPE--NSFAQAFDAANIITL 231
DB 175 YFDIQELFFRTVDSATEFLFESVHSLRDEEIGYDTKDMAEERRKPADAFNKSQVILST 234
QY 232 RFIDPLMRKFRFFHVGSEALLAQSIKLVDFT--YSVIRRRKAEIIVEVRASG-----281
DB 235 RV-----ALQTLVNLVNNKEFCNDIVHKFTNYVQKALDTPBELEKQGGVFLYELA 289
QY 282 KQEKMKHILSRFIELGEAGDGGGDKSLRDVVLNFIAGRDTTATTLTTSWFTHMAMS 341
DB 290 KQTR-----DPKVLDDQSLNILLAGRDTTAGLLSFAVFEELAR 326
QY 342 HPDVAEKLRLRELCAPAEAREBEGVTLVLCGGADADDKAFARVAQFAGLLTYDSLGLV 401
DB 327 NPHIWAQLREEI-----ESHFGL-----GED-----SRVEE-----ITFESLKRCE 362
QY 402 YLHACVTTETRLYPVAVPQDPKGILEDDVLPDGTQKVRAGG-----MVTYVPYSMGR 451
DB 363 YLKAVLNELTLRHPSVPRNARPAIKDTTLP-----RGGPGNGKDPILIRKDEVVQYSISA 417
QY 452 MEYN---WGPDAASFRPERWINEGDAFRNASPFKFTAFQAGPRICLGKDSAYLQMKWALA 508
DB 418 TQTNPAVYAGDAADFRPERWF--EPSTRNLG-WAFLPFGNGGPRICLGQGFALTEAGYVLV 474
QY 509 ILFRFYSFRLLEGHP 523
DB 475 -----RLVQEF 481

```

```

RESULT 12
US-09-302-620B-99
; Sequence 99, Application US/09302620B
; Patent No. 6331420
; GENERAL INFORMATION:
; APPLICANT: Willson, C. Ron
; APPLICANT: Craft, David L.
; APPLICANT: Birch, Dudley

```

APPLICANT: Bshoo, Mark  
APPLICANT: Madduri, Krishna M.  
APPLICANT: Cornett, Cathy A.  
APPLICANT: Brenner, Alfred A.  
APPLICANT: Tang, Maria  
APPLICANT: Loper, John C.  
APPLICANT: Gleeson, Martin  
TITLE OF INVENTION: CYTOCHROME P450 MONOOXYGENASE AND NADPH CYTOCHROME P450  
TITLE OF INVENTION: OXIDOREDUCTASE GENES AND PROTEINS RELATED TO THE OMEGA  
TITLE OF INVENTION: HYDROXYLASE COMPLEX OF CANDIDA TROPICALIS AND METHODS  
TITLE OF INVENTION: RELATING THERETO  
FILE REFERENCE: 1010-16:seq  
CURRENT APPLICATION NUMBER: US/09/302,620B  
CURRENT FILING DATE: 1999-04-30  
NUMBER OF SEQ ID NOS: 109  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 99  
LENGTH: 540  
TYPE: PRT  
ORGANISM: Candida tropicalis  
US-09-302-620B-99

Query Match 12.3%; Score 379; DB 4; Length 540;  
Best Local Similarity 26.4%; Pred. No. 4.le-33;  
Matches 126; Conservative 74; Mismatches 187; Indels 90; Gaps 14;  
QY 97 DPVNVHVLKTNFTNPKGIYRSYMDVLLGDGIFNADGELMRKQRTASFEFASXNLRD 156  
Db 122 DPENIKAVLATQNFDFSLG-TRHDFLYSLGDDGIFTDGAGWKHSRMTLRPQFAREQVSH 180  
QY 157 FSAI-----VPREYSLKLSGLSQAASKAGKVVDMQELMYMTLDSICKVGFVEIGTLP 211  
Db 181 VKLLEPHVQVFFKHVRK-----HRGQTDFDIQELFRLLTVDSATEFLLFGSABSLRD 231  
QY 212 DL-----PENSEFAAFAAIIITLRFIDPLWRIKRPFHVGEALLAQSIKLV 260  
Db 232 DSVGLTPTTDFEGRGDFADAFNSQTYQARFL-----LQQMWILNGAEFRKSIALVH 286  
QY 261 EFTYSVIRRAKAEIVVEVRASQKQKMDHLSRFLIEGEAGDDGGGDKSLRDVVVNF 320  
Db 287 KFADHYV-QKALELTD-----DLQKQGVVFLYELAKQ-----TRDPKVLRLDQLNI 333  
QY 321 VIAGRDTATTLTSPHWMHSHPDVAEKLRELCFAFAERAREBGVTILVLCGGADADKA 380  
Db 334 LVAGRDTATTLTSPHWMHSHPDVAEKLRELCFAFAERAREBGVTILVLCGGADADKA 380  
QY 381 FA-----ARVAQFAGLLTVSLGKLVYLHACVTEITRLYPAYPQDPKGILEDDVLPD 433  
Db 370 FGLGEEARVES-----ISFESLKSCEYLKAVINEALRLYPSVPHNFRVATRNITLPRGGK 425  
QY 434 -----TKVRAGMVTYVPYSMGRMEYNWGPDAASFRPERWINEGAFNAGSPFKETAPOA 488  
Db 426 DGCSPVIVKGVVMTYVIGTHRDPISYICADADVFRPERWFEPE-----TRKLGWAYVPFNG 482  
QY 489 GPRICLGKDSAYLQMKMALILFRFYSFRLLE-----GHPVQVYRMVMTILSWAGLKVRV 542  
Db 483 GPRICLGQQAFTASVYTVRLQLQEFGLNLSLDPNAEYPPKLQNTLITLSLFDGADVRM 539

RESULT 13  
US-08-948-564-16  
Sequence 16, Application US/08948564  
Patent No. 6121512  
GENERAL INFORMATION:  
APPLICANT: Siminszky, Balazs  
APPLICANT: Dewey, Ralph E.  
APPLICANT: Corbin, Frederick T.  
TITLE OF INVENTION: No. 6121512el Cytochrome P-450 Constructs and  
TITLE OF INVENTION: Methods of Producing Herbicide-Resistant Transgenic Plants  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Virginia C. Bennett  
STREET: PO Box 37428

CITY: Raleigh  
STATE: No. 6121512th Carolina  
COUNTRY: USA  
ZIP: 27627  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/948,564  
FILING DATE:  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Bennett, Virginia C.  
REGISTRATION NUMBER: 37,092  
REFERENCE/DOCKET NUMBER: 5051-409  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-854-1400  
TELEFAX: 919-854-1401  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 576 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-948-564-16

Query Match 11.9%; Score 367; DB 3; Length 576;  
Best Local Similarity 24.3%; Pred. No. 1e-31;  
Matches 142; Conservative 96; Mismatches 222; Indels 124; Gaps 24;  
QY 16 PF---PFLAGPKYIALLLVLSWILVQWLSLRFKQKQPR-----S 52  
Db 27 PFTHFSLSQPKR-----ISSIRCSQINTDKKSSRNLLGNASNLLTDLSSGSGS 78  
QY 53 WPVIGATVEQLRN---YHRMHDWLVGYLSHRHTVTVDMPFTSYTYIADPVNVEHLKTNF 109  
Db 79 MPAEAGNVDLLGRPLFFSLYDW---FLEHGAVVYKLAFGPKAFVWVSDPIVARHILRENA 135  
QY 110 TNPYKGIYRSYMDVLLGDGIFNADGELMRKQRTASFEFASXNLRDFFSAIVREYS--- 166  
Db 136 FSYDKGVL-ADILEPIMGKGLIPADLDTMKQRRRVIAFAPHNSYL-EAMVKIITCSERT 193  
QY 167 -LKLSGLSQAASKAGK---VDMQELMYMTLDSICKVGFVEIGTLPDLP--ENSFAQ 220  
Db 194 ILKFNKLEGEVGDGDPDSIELDLAEFSSALDILGLGVENYDFGVTKSPVYKAVYGT 253  
QY 221 AFDAANIITLRFIDPLWRI--KFFHVGSEALLAQSIKLVDEFTYSVIRRRKAEIVEVR 278  
Db 254 LFEAEH---RSTFYIPYKIPLARWI-VPRQRKFQDDLKVINTCLDGLIRNAKESRQETD 309  
QY 279 ASGQKEK-----MKHDIUSRFLIEGEAGDDGGGDKSLRDVILNFIAGRDITATLSW 334  
Db 310 VEKLQQRDYLNLKDALSLRLFLV---DMRGADVDDRLQRLDMLTMLIAGHETFAAVLTW 364  
QY 335 FTHWAMSHPDVAEKLRELCFAFAERAREBGVTILVLCGGADADKAFAARVAQFAGLLTY 394  
Db 365 AVFLLAQNSKMKKAQAE-----VDLVLTGRP-----TF 394  
QY 395 DSLGKLVYLHACVTEITRLYPAYPQDPKGILEDDVLP-----DGTVKAGGMVTVVPY 447  
Db 395 ESLKELQYIRLIVVEALRLYPPQLLIRSLKSDVLPFGHKGEKGDGYAIPAGTDFVTSVY 454  
QY 448 SMGRMEYNWG-PDAASFRPERWI---NED-----GAFRNASP-----FKETAF 486  
Db 455 NLHRSYFWDKRPD--DPEPERFLVQNKNEELEGWAGLDPRSGPGLAYPNEVISDFALPFP 512  
QY 487 QAGPRICLGKDSAYLQMKMALILFRFYSFRLLEGHGPVQYRMVT 530  
Db 513 GGGPRKCVGQDFALMESTVAUTMLLQNFVBE-LKGTPESELVLT 555

```
RESULT 14
US-09-302-620B-104
; Sequence 104, Application US/09302620B
; Patent No. 6331420
; GENERAL INFORMATION:
; APPLICANT: Wilson, C. Ron
; APPLICANT: Craft, David L.
; APPLICANT: Birch, Dudley
; APPLICANT: Eshoo, Mark
; APPLICANT: Madduri, Krishna M.
; APPLICANT: Cornett, Cathy A.
; APPLICANT: Brenner, Alfred A.
; APPLICANT: Tang, Maria
; APPLICANT: Loper, John C.
; APPLICANT: Gleeson, Martin
; TITLE OF INVENTION: CYTOCHROME P450 MONOOXYGENASE AND NADPH CYTOCHROME P450
; TITLE OF INVENTION: OXIDOREDUCTASE GENES AND PROTEINS RELATED TO THE OMEGA
; TITLE OF INVENTION: HYDROXYLASE COMPLEX OF CANDIDA TROPICALIS AND METHODS
; TITLE OF INVENTION: RELATING THERETO
; FILE REFERENCE: 1010-16. seq
; CURRENT APPLICATION NUMBER: US/09/302,620B
; CURRENT FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 109
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 104
; LENGTH: 499
; TYPE: PRT
; ORGANISM: Candida tropicalis
US-09-302-620B-104

Query Match 11.5%; Score 353; DB 4; Length 499;
Best Local Similarity 26.3%; Pred. No. 2.9e-30;
Matches 127; Conservative 79; Mismatches 182; Indels 94; Gaps 17;

QY 80 HRTVDMFTSYTYIADPNVEHLKTNFTNPKGIVYRSMVMDVLLGDGIFNADGELWK 139
DB NRALSVEPILTK-----DPVNIKAMJSTQDDPSLGLRLHQFAP-LLGKGIFFLDGPEWK 142
QY 140 KQKKTASFAS-----KNIKRDFA--IVFREYSIKLSGILSQASKAGKGVMDQBELYMRMTL 194
DB QSRSMRLPQFAKQVRSHLDLPHFVLLRGH---IDG-----HNGDYFDIQELYFRFSM 193
QY 195 DSICKVGFVEIGTLPDLPENSAQAFDAANIITLRFIDPLWRIKRFHVGSEALLAQ 254
DB DVATGELFGESVGLKDE--DARFLAFNESQKYLATR-----ATLHE 234
QY 255 SIKLVDEFTY-----SVIRREKAEIVRASGKQEKMKHDILSRFIELGEAGDDGGGFGDD 310
DB LYFLCDGFRQYKVKRKCQCVH-----KALDVAPETSYSYVFLRELVKH---TRDP 286
QY 311 KSLRDVVLNFVIAGRDPTATTLNFWTHMAMSHDPVAEKLRLREICAFEAERAEETVLVL 370
DB VVLQDQALVLLAGRDUTTSLSPATPELARNDHMRKREE-----VIL 331
QY 371 CGGADDDKFAARVAQFAGLLTYDSLGKLVYLHACVTETRLYPVPQDPKGILEDVVL 430
DB TMGPSSDE-----ITVAGLKSCRYLKAILENETRLYPSPVRNARFATRNITL 378
QY 431 PDG-----TKVRAGMVTYVYPSMGRMEYNWGPDAASFPRPRWINEGDAFNASPFK 482
DB PRGGPDGSPFILIRKQPGVGYFCATHLNEKVGNDSHVFRPRWAALEG---XSLGWS 435
QY 483 FTAFQAGPRICLGKDSAYLQMMALAILFRPYS---FLLLEGHPVQVYRMMTILSMAGLK 539
DB YLPNGGSPICLGQGFALLESAYVLARLTQCYTTIQLRTTEYPPKLVHLT-MSLLNGVY 494
QY 540 VR 541
DB 495 IR 496
```

RESULT 15  
US-09-976-594-533

```
; Sequence 533, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 533
; LENGTH: 524
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 6825202CD1
US-09-976-594-533

Query Match 11.5%; Score 353; DB 4; Length 524;
Best Local Similarity 23.8%; Pred. No. 3.2e-30;
Matches 138; Conservative 97; Mismatches 226; Indels 118; Gaps 19;

QY 9 LTPATPSPPFPLAGPHKXIALLLVLSWILYQKSLRQKGRSPWPVIGATVEQLRNYHR 68
DB LGPVAASP-----WLLLLVGGSWLLAR-----VLAWTYTFYDNCRR 47
QY 69 M-----HDLVGYL-----SRHRTVTVDMPFTSY-----TYIADPVNV 101
DB LQCFPPQPKQNFQWGHQGLVPTTEGMMKTLTQLVTTYQGFKLWLGPTFPLLLCHPDII 107
QY 102 EHVLTNFTNTPKGVYRSMYMDVLLGDGIFNADGELWKQKRTASFEFASKNLRDFSIV 161
DB RPTSASAAVAPKDMIFYGFLKPLWLDGILLSGGKWSRHRRLMTPAFHFNILKPYMKIF 167
QY 162 FREYSLKLSGILSQASKAGKGVMDQBELYMRMTLDSICKVGFVEIGT-----LSPDLP 214
DB NKSVMIMHDKWRLASGSAKLDMPFHLISLMTLDSLQKCVSFESNCOEKPSYIAALIE 227
QY 215 ENSFAQAFDAANIIT--LRFIDP-LWRIKRFHVGSEALLAQSIKLVDEFYTSVIRRK 271
DB LSFAVEKRNQQLLHHTDFLYITPDGQRRACH-----LVHDFDVAVIQERR 275
QY 272 ABIVEVRASGKQEKMKHDILSRFIELGE-----AGDDGGFGDDKSLRDVVLNFIAGRT 327
DB RTL---PTQIDDFLKNKAKSKTLDIFDVLVLLSKDDEKELSDDEDIRABADTFMEGHDT 332
QY 328 TATLISWFTMHAMSHDPVAEKLRLREICAFEAERAEETVLVLGGADADDKAFARVAQ 387
DB TASGLSWVLVHLAKHPEYQECRQEVQELLKQREPIE----- 369
QY 388 FAGLLTYDSLGKLVYLHACVTETRLYPVPQDPKGILEDVVLPGTKVRAAGMVTYVPY 447
DB -----IEWDLDLAQLPFTLWIKESLRLHPPVVISRCTQDFVLPDG-RVIPKGIIVCLI-- 422
QY 448 SMGRMEYN---MGPDAAFRPRWINEGDAFNASPFKTAQAGPRICLGKDSAYLQMK 504
DB NIIGHYNTVW-PDPEVYDPPFRDQEN--IKERSPLAFIPFSAGPRNCIGQAFAMAEMK 479
QY 505 MALAILFRPYSFLLLEGHPVQVYRMMTILSMH-GLKVRV 542
DB VLVAL--TLLHFRILPTHTPRKKPELILRAEGGLIWRV 516

Search completed: March 2, 2004, 15:30:19
Job time : 24 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2004, 15:21:34 ; Search time 60 Seconds  
(without alignments)  
2768.966 Million cell updates/sec

Title: US-10-021-657-2

Perfect score: 3079

Sequence: 1 EFGTREAHLPATPPSPFFPL.....ICMHACKGRKRWVSLVAWLKP 588

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A Geneseq\_29Jan04.\*

1: Geneseqp1980s.\*

2: Geneseqp1990s.\*

3: Geneseqp2000s.\*

4: Geneseqp2001s.\*

5: Geneseqp2002s.\*

6: Geneseqp2003as.\*

7: Geneseqp2003bs.\*

8: Geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3079	100.0	588	5	ABB08037
2	976	31.7	541	2	AA01648
3	976	31.7	541	2	AA01647
4	971.5	31.6	537	3	AA01647 Cytochrom
5	970.5	31.5	530	3	AA01647 Cytochrom
6	969.5	31.5	523	3	AA01647 Cytochrom
7	961.5	31.2	522	3	AA01647 Cytochrom
8	961.5	31.2	527	3	AA01647 Cytochrom
9	911.5	29.6	489	3	AA01647 Cytochrom
10	894	29.0	475	3	AA01647 Cytochrom
11	886	28.8	475	3	AA01647 Cytochrom
12	884	28.7	465	3	AA01647 Cytochrom
13	847	27.5	508	3	AA01647 Cytochrom
14	847	27.5	510	3	AA01647 Cytochrom
15	847	27.5	528	3	AA01647 Cytochrom
16	841	27.3	508	3	AA01647 Cytochrom
17	841	27.3	510	3	AA01647 Cytochrom
18	841	27.3	527	3	AA01647 Cytochrom
19	840	27.3	512	2	AA01647 Cytochrom
20	832	27.0	511	2	AA01647 Cytochrom
21	831.5	27.0	510	2	AA01647 Cytochrom
22	818.5	26.6	513	2	AA01647 Cytochrom
23	802.5	26.1	511	2	AA01647 Cytochrom
24	785.5	25.5	514	2	AA01647 Cytochrom
25	780	25.3	519	5	ABB90978

26	773.5	25.1	477	3	AA023271	Arabidops
27	770.5	25.0	448	3	AA023272	Arabidops
28	770	25.0	158	5	ABB08038	Sorghum S
29	760.5	24.7	464	6	ADA48304	Rice proc
30	747.5	24.3	438	3	AA023273	Arabidops
31	744	24.2	516	3	AA023273	Arabidops
32	738.5	24.0	510	3	AA023273	Arabidops
33	730	23.7	479	3	AA023273	Arabidops
34	728	23.6	475	3	AA023273	Arabidops
35	727.5	23.6	506	3	AA023273	Arabidops
36	726.5	23.6	437	3	AA023273	Arabidops
37	721.5	23.4	495	3	AA023273	Arabidops
38	721.5	23.4	497	3	AA023273	Arabidops
39	721.5	23.4	507	3	AA023273	Arabidops
40	720	23.4	441	3	AA023273	Arabidops
41	693.5	22.5	464	3	AA023273	Arabidops
42	693.5	22.5	471	3	AA023273	Arabidops
43	687	22.3	428	3	AA023273	Arabidops
44	685	22.2	463	3	AA023273	Arabidops
45	685	22.2	467	3	AA023273	Arabidops

## ALIGNMENTS

### RESULT 1

ABB08037  
ID ABB08037 standard; protein; 588 AA.

XX ABB08037;

DT 27-AUG-2002 (first entry)

XX Maize SEMu200 protein.

XX SEMu200; fertility; maize; hybrid seed; sterility; plant; transgenic.

XX Zea mays.

XX Key Location/Qualifiers

FT Region 1..546

FT /note= "N-terminal portion of SEMu200"

FT Region 546..547

FT /note= "point of fusion between the N- and C- terminal

FT parts of SEMu200, separated by a stop codon in the

FT polynucleotide sequence"

FT Region 547..588

FT /note= "C-terminal portion of SEMu200"

XX WO200226789-A2.

XX 04-APR-2002.

XX 25-SEP-2001; 2001WO-US029886.

XX 26-SEP-2000; 2000US-00670153.

XX (PTON-) PIONEER HI-BRED INT INC.

XX Albertsen MC, Fox T, Huffman G, Trimmell M;

XX WPI; 2002-471258/50.

XX N-PSDB; ABL60611.

XX Novel nucleic acid sequence comprising SEMu200 gene, useful for producing  
transgenic plants and mediating male fertility in plants.  
XX Claim 2; Fig 4; 60pp; English.  
XX The invention relates to a SEMu200 gene, useful for impacting male  
fertility in a plant e.g. maize. An isolated nucleotide sequence  
comprising the SEMu200 gene is useful for producing hybrid seeds and for  
providing heritable externally controllable male sterility in a plant.

CC The present sequence represents the maize SBW1200 protein

XX  
SQ Sequence 588 AA;  
Query Match 100.0%; Score 3079; DB 5; Length 588;  
Best Local Similarity 100.0%; Pred. No. 1.8e-303;  
Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 EFGTREAHLTPATPSPPFFLAGPHKVIALLVLSWLVQWLSRKKQKQKPRSWPVIGATV 60  
DB 1 EFGTREAHLTPATPSPPFFLAGPHKVIALLVLSWLVQWLSRKKQKQKPRSWPVIGATV 60  
QY 61 EQLRNVRHMDWLVGLVLSRHRITVDMPTSTYTIADPNVVEHLKTNFTNPKGIYVRS 120  
DB 61 EQLRNVRHMDWLVGLVLSRHRITVDMPTSTYTIADPNVVEHLKTNFTNPKGIYVRS 120  
QY 121 YMDVLGDIENAGBELWKRKORTASFEFASKNLRDPSAIVREYSKLGLSLQASKAG 180  
DB 121 YMDVLGDIENAGBELWKRKORTASFEFASKNLRDPSAIVREYSKLGLSLQASKAG 180  
QY 181 KVVDMOELVWRMTLDSICKVGFVEIGTISPDLPENSFAQAFDAANIITLRFIDPLWRI 240  
DB 181 KVVDMOELVWRMTLDSICKVGFVEIGTISPDLPENSFAQAFDAANIITLRFIDPLWRI 240  
QY 241 KRFFHYGSEALAQSIKLVDEFTYVIRRRKABIEVEVRASGKQKMKHDLISRFELGEA 300  
DB 241 KRFFHYGSEALAQSIKLVDEFTYVIRRRKABIEVEVRASGKQKMKHDLISRFELGEA 300  
QY 301 GDDGGGDDKSLRDVVLNFIAGRTTATLTSWETHMAMSHPDVAEKLRRRELCAFEAE 360  
DB 301 GDDGGGDDKSLRDVVLNFIAGRTTATLTSWETHMAMSHPDVAEKLRRRELCAFEAE 360  
QY 361 AREEGVTLVLCGGADADDKAFARVAQFAGLLTYDSLGKLVYLHACVTETLRLYPVAPQD 420  
DB 361 AREEGVTLVLCGGADADDKAFARVAQFAGLLTYDSLGKLVYLHACVTETLRLYPVAPQD 420  
QY 421 PKGILEDDVLPDGTQVRAGMVTVPYSGRMMEYNWGPDAASPRPERWINEDEGAFRNASP 480  
DB 421 PKGILEDDVLPDGTQVRAGMVTVPYSGRMMEYNWGPDAASPRPERWINEDEGAFRNASP 480  
QY 481 FKFTAFQAGPRICLGKDSAYLQMKWALAILFRFYSFRLLEGGHPVQYRMMTILSMAHGLKV 540  
DB 481 FKFTAFQAGPRICLGKDSAYLQMKWALAILFRFYSFRLLEGGHPVQYRMMTILSMAHGLKV 540  
QY 541 RVSAVCHGDLMDIVPLNPRQITLVLIQICMHWACKGRWVSLVAVLKP 588  
DB 541 RVSAVCHGDLMDIVPLNPRQITLVLIQICMHWACKGRWVSLVAVLKP 588

RESULT 2  
ID AAY01648  
AA Y01648 standard; protein; 541 AA.

XX AC AAY01648;  
XX DT 23-JUN-1999 (first entry)  
XX DE Cytochrome P450 encoded by a codon modified cDNA.  
XX DE Cytochrome P450; protein expression; yeast; mutant.  
XX OS Synthetic.  
XX OS Triticum aestivum.  
XX PN FR2768748-A1.  
XX PD 26-MAR-1999.  
XX PF 24-SEP-1997; 97FR-00012094.  
XX PR 24-SEP-1997; 97FR-00012094.  
XX PA (RHON ) RHONE-FOULENC AGROCHIMIE.

XX Bataud Y, Schalk M, Durst P, Werck RD;  
XX WPI; 1999-217499/19.  
XX N-PSDB; AAX26878.  
XX DNA for expression in yeasts - with codon changes based on yeast codon usage.  
XX Example 4; Page 24-26; 31pp; French.  
XX The present sequence represents a cytochrome P450, encoded for by a cDNA which was created by altering the codons of AAX26865 to exemplify the invention. The specification describes a DNA sequence that encodes a protein of interest and contains regions with a high content of codons poorly suited to yeasts, where a sufficient number of these codons are replaced by corresponding codons suited to yeasts in these regions.  
XX Yeasts transformed with vectors such DNA sequences can be cultured to produce the protein of interest (especially a plant protein) or, when the protein is an enzyme (such as cytochrome P450), can be cultured in the presence of a substrate for the enzyme to produce a substrate conversion product  
XX Sequence 541 AA;  
Query Match 31.7%; Score 976; DB 2; Length 541;  
Best Local Similarity 39.9%; Pred. No. 1.2e-89;  
Matches 222; Conservative 82; Mismatches 190; Indels 62; Gaps 13;  
QY 27 IALLVLSWLVQWLSL-RKQKPRSWPVIGATVEQLRNVRHMDWLVGLYLSR---HR 81  
DB 10 VSAAVAAWAF---WRMSRGLRGRVWVPLGSLPLGVQHAEDMHEWIAGNLRAGGTQ 65  
QY 82 TVTVDMPFTS---YTYIADPNVVEHLKTNFTNPKGIYVRSYMDVLGGLGFINADGE 136  
DB 66 TCIFAVGVARRGGLVTVCPRNLEHLVKARFDNYPKGFPHVGVFRDILGDIENSDGD 125  
QY 137 LWRKOKRTASFEFASKNLRD-FSAIVREYSKLGLSLQASKAGKVVDMOELVWRMTLD 195  
DB 126 TWLAQKRTAALFTTTLRTATMSRWVRSIHGRLPLILADAAGKQAVDQLDILLRLTFD 185  
QY 196 SICKVGFVEIGTISPDLPENSFAQAFDAANIITLRFIDP-LWRPKRFFHYGSEALLA 253  
DB 186 NICGLAFGKDPETLAQGLPENEFASAFDRATEATLNRFFPEBFLWCKKWLGLMETTILT 245  
QY 254 QSTKLVDEFTYVIRRRKABIEVEVRASGKQKMKHDLISRFELGAGDDGGFGDDKSL 313  
DB 246 SSMHVVDQVLAIVIKRKLAAAGNGKCDTAAATHDILLSRFMRKGSYSDE-----SL 297  
QY 314 RDVVLNFIAGRTTATLTSWETHMAMSHPDVAEKLRRRELCAFEAEAREEGVTLVLCGG 373  
DB 298 QHVALNFIAGRTISSVALSWFFWLVSTHPAVERKIVRELCS-----VLAAS 344  
QY 374 ADADKAFARVAQFAGLLTYDSLGKLVYLHACVTETLRLYPVAPDPPKGILEDDVLPD 433  
DB 345 RGAHPAL-----WLAEPFTFEELDRILVYKALSETLRLYPSVPEDSKHVVADYLPD 399  
QY 434 TKVRAGMVTVPYSGRMMEYNWGPDAASPRPERWINEDEGA-FRNASPKFTAFQAGPRI 492  
DB 400 TFPAGSSVTYISYAGRMKGVWGEDCLFEPFRWRWSADGTFKEQHDSDYKFAVFNAGPRV 459  
QY 493 CLGKDSAYLQMKWALAILFRFYSFRLLEGGHPVQYRMMTILSMAHGLKVRVSRVACHGDL 551  
DB 460 CLGKDLAYLQMKWALAILFRFYSFRLLEGGHPVQYRMMTILSMAHGLKVRVSRVACHGDL 551  
QY 552 DMDIVPLNPRQITLV 567  
DB 510 -----RPRDLAPVL 518  
RESULT 3  
ID AAY01647  
AA Y01647 standard; protein; 541 AA.

XX	AAV01647;	Db	345	RGADPAL-----WLAEPFTFEELDRVLKALSETLRLYSPVDESKHVVDYDLPDG	399
AC					
XX	23-JUN-1999 (first entry)	Qy	434	TKVRAGGMVTVPYISMGRMEYNWGPDAASFRPERWINEDEGA-FRNASPKFTAFQAGPRI	492
XX					
DE	Cytochrome P450 designated CYP86A5.	Db	400	TFVPAGSSVTYSIYSAGRMKGVMGEDCCLFPRPERWLSADGTFKFEQDHSYKFFVAFNAGPRV	459
XX					
KW	Cytochrome P450; protein expression; yeast.	Qy	493	CLGKDSAYLQMK-MALAJLFRFYSERLLEGHFVQVRMTTILSMHGLKVRVSRVAVCHGDL	551
XX					
OS	Triticum aestivum.	Db	460	CLGKDLAYLQMKNIAGSVLLR-HRLTVAPGHRVQKMSLTLMKGLRMEV-----	509
XX					
XX	FR2768748-A1.	Qy	552	DMDIVELNPRQITLVL	567
XX					
XX	26-MAR-1999.	Db	510	-----RPRDLAPVL	518
XX					
XX	24-SEP-1997; 97PR-00012094.				
XX					
XX	24-SEP-1997; 97PR-00012094.				
XX	(RHON ) RHONE-POULENC AGROCHIMIE.				
XX					
PI	Bataard Y, Schalk M, Durst F, Werck RD;				
XX					
XX	WPI; 1999-217499/19.				
DR	N-PSDB; AAX26874.				
XX					
XX	DNA for expression in yeasts - with codon changes based on yeast codon				
PT	usage.				
XX					
XX	Example 4; Page 21-23; 31pp; French.				
XX					
CC	The present sequence represents a cytochrome P450, and was used to				
CC	exemplify the invention. The specification describes a DNA sequence that				
CC	encodes a protein of interest and contains regions with a high content of				
CC	codons poorly suited to yeasts, where a sufficient number of these codons				
CC	are replaced by corresponding codons suited to yeasts in these regions.				
CC	Yeasts transformed with vectors such DNA sequences can be cultured to				
CC	produce the protein of interest (especially a plant protein) or, when the				
CC	protein is an enzyme (such as cytochrome P450), can be cultured in the				
CC	presence of a substrate for the enzyme to produce a substrate conversion				
CC	product				
XX					
XX	Sequence 541 AA;				
XX					
XX	Query Match				
XX	Best Local Similarity 31.7%; Score 976; DB 2; Length 541;				
XX	Matches 222; Conservative 82; Mismatches 190; Indels 62; Gaps 13;				
Qy	27 IALLVLSWILVQWMSL-RKQKGRSPWPVIGATVEQLNVRHMDLWLVLSR-----HR	81			
Db	10 VSAAVAAVMAWF-----WRMSRGLRGPVWPVLGSLFGLVQHAEDMHEWLAGNLRRAAGGTIQ	65			
Qy	82 TVTVDMPTTS-----YTVIADPVNVEHVLKNTFTYKPGIVVRSYMDVLLGDGIFENADGE	136			
Db	66 TCIFAVGVARGGLVTVTCDPNLEHLVKARFDNYPKPGFWHGFVDRDLGDGIFNSDGD	125			
Qy	137 LMRKORKTASFFASKNLRD-FSAIVFREYSKLSGLISQASKAGKVDMDQBLYNRMILD	195			
Db	126 TWLAQRKTAALFTTTLRTAMSRVSRSHGRLLPILADAAGAQAQVDQLDLRLTFD	185			
Qy	196 STCKVGFVGEITLSPDLPENSAQAFDAANIITLRFIDP--LWRIKFFHVGSEALLA	253			
Db	186 NICGLAFGKDPETLAQGLPENEFASAFDRATEATLNRFFPFBLMRCKKWLGLMETTUT	245			
Qy	254 QSIKLVDETYSVIRRKAEIIEVVRASQKQKMKHDIILSRFTELGEAGDGGGFGDKSL	313			
Db	246 SSHAHVDDYLAIVIKRKKLELAGNCKDCTAATHDLLSRFRKGSYSDE-----SL	297			
Qy	314 RDVLNFIAGRDTTATLTSWFTTHAMSHPDVVAEKLRELCAFEAPAREEGVTILVLCGG	373			
Db	298 QHVALNFIAGRDTSSVALSWFFWLVSTHFAVERKIVRELCS-----VLAAS	344			
Qy	374 ADADKAFARVAQFAGLLTYSGLKVLVHACVTETLRLYPVAVPQDPKGILEDVLPDG	433			

us-10-021-657-2.rag

Mon Mar 8 10:13:27 2004

PR	01-JUN-1999;	99US-0137222P.	PR	05-AUG-1999;	99US-0147192P.
PR	03-JUN-1999;	99US-0137528P.	PR	05-AUG-1999;	99US-0147260P.
PR	04-JUN-1999;	99US-0137502P.	PR	06-AUG-1999;	99US-0147303P.
PR	07-JUN-1999;	99US-0137724P.	PR	06-AUG-1999;	99US-0147416P.
PR	08-JUN-1999;	99US-0138094P.	PR	09-AUG-1999;	99US-0147933P.
PR	10-JUN-1999;	99US-0138540P.	PR	09-AUG-1999;	99US-0147935P.
PR	10-JUN-1999;	99US-0138847P.	PR	10-AUG-1999;	99US-0148171P.
PR	14-JUN-1999;	99US-0139119P.	PR	10-AUG-1999;	99US-0148319P.
PR	16-JUN-1999;	99US-0139453P.	PR	11-AUG-1999;	99US-0148341P.
PR	16-JUN-1999;	99US-0139453P.	PR	12-AUG-1999;	99US-0148565P.
PR	17-JUN-1999;	99US-0139492P.	PR	13-AUG-1999;	99US-0148684P.
PR	18-JUN-1999;	99US-0139454P.	PR	13-AUG-1999;	99US-0149368P.
PR	18-JUN-1999;	99US-0139455P.	PR	16-AUG-1999;	99US-0149175P.
PR	18-JUN-1999;	99US-0139456P.	PR	17-AUG-1999;	99US-0149175P.
PR	18-JUN-1999;	99US-0139457P.	PR	18-AUG-1999;	99US-0149426P.
PR	18-JUN-1999;	99US-0139458P.	PR	20-AUG-1999;	99US-0149723P.
PR	18-JUN-1999;	99US-0139459P.	PR	20-AUG-1999;	99US-0149723P.
PR	18-JUN-1999;	99US-0139460P.	PR	20-AUG-1999;	99US-0149929P.
PR	18-JUN-1999;	99US-0139461P.	PR	23-AUG-1999;	99US-0149902P.
PR	18-JUN-1999;	99US-0139462P.	PR	23-AUG-1999;	99US-0149930P.
PR	18-JUN-1999;	99US-0139463P.	PR	25-AUG-1999;	99US-0150566P.
PR	18-JUN-1999;	99US-0139750P.	PR	25-AUG-1999;	99US-0150884P.
PR	18-JUN-1999;	99US-0139763P.	PR	26-AUG-1999;	99US-0151065P.
PR	21-JUN-1999;	99US-0139817P.	PR	27-AUG-1999;	99US-0151066P.
PR	22-JUN-1999;	99US-0139899P.	PR	27-AUG-1999;	99US-0151080P.
PR	23-JUN-1999;	99US-0140353P.	PR	30-AUG-1999;	99US-0151303P.
PR	23-JUN-1999;	99US-0140354P.	PR	30-AUG-1999;	99US-0151303P.
PR	24-JUN-1999;	99US-0140695P.	PR	31-AUG-1999;	99US-0151438P.
PR	28-JUN-1999;	99US-0140823P.	PR	01-SEP-1999;	99US-0151930P.
PR	29-JUN-1999;	99US-0140891P.	PR	01-SEP-1999;	99US-0152363P.
PR	30-JUN-1999;	99US-0141287P.	PR	07-SEP-1999;	99US-0153070P.
PR	01-JUL-1999;	99US-0141842P.	PR	10-SEP-1999;	99US-0153758P.
PR	01-JUL-1999;	99US-0142154P.	PR	13-SEP-1999;	99US-0154018P.
PR	02-JUL-1999;	99US-0142055P.	PR	15-SEP-1999;	99US-0154039P.
PR	06-JUL-1999;	99US-0142390P.	PR	20-SEP-1999;	99US-0154779P.
PR	08-JUL-1999;	99US-0142803P.	PR	22-SEP-1999;	99US-0155139P.
PR	09-JUL-1999;	99US-0142920P.	PR	23-SEP-1999;	99US-0155486P.
PR	12-JUL-1999;	99US-0142977P.	PR	24-SEP-1999;	99US-0155659P.
PR	13-JUL-1999;	99US-0143542P.	PR	28-SEP-1999;	99US-0156458P.
PR	14-JUL-1999;	99US-0143624P.	PR	29-SEP-1999;	99US-0156596P.
PR	15-JUL-1999;	99US-0144005P.	PR	04-OCT-1999;	99US-0157117P.
PR	16-JUL-1999;	99US-0144086P.	PR	05-OCT-1999;	99US-0157753P.
PR	16-JUL-1999;	99US-0144325P.	PR	06-OCT-1999;	99US-0157865P.
PR	19-JUL-1999;	99US-0144331P.	PR	07-OCT-1999;	99US-0158029P.
PR	19-JUL-1999;	99US-0144332P.	PR	08-OCT-1999;	99US-0158232P.
PR	19-JUL-1999;	99US-0144333P.	PR	12-OCT-1999;	99US-0158369P.
PR	19-JUL-1999;	99US-0144334P.	PR	13-OCT-1999;	99US-0159293P.
PR	19-JUL-1999;	99US-0144335P.	PR	13-OCT-1999;	99US-0159294P.
PR	19-JUL-1999;	99US-0144336P.	PR	13-OCT-1999;	99US-0159295P.
PR	19-JUL-1999;	99US-0144337P.	PR	14-OCT-1999;	99US-0159329P.
PR	19-JUL-1999;	99US-0144338P.	PR	14-OCT-1999;	99US-0159330P.
PR	19-JUL-1999;	99US-0144339P.	PR	14-OCT-1999;	99US-0159331P.
PR	20-JUL-1999;	99US-0144884P.	PR	14-OCT-1999;	99US-0159637P.
PR	21-JUL-1999;	99US-0144814P.	PR	14-OCT-1999;	99US-0159638P.
PR	21-JUL-1999;	99US-0145086P.	PR	18-OCT-1999;	99US-0159584P.
PR	21-JUL-1999;	99US-0145088P.	PR	21-OCT-1999;	99US-0160741P.
PR	22-JUL-1999;	99US-0145085P.	PR	21-OCT-1999;	99US-0160767P.
PR	22-JUL-1999;	99US-0145087P.	PR	21-OCT-1999;	99US-0160768P.
PR	22-JUL-1999;	99US-0145089P.	PR	21-OCT-1999;	99US-0160770P.
PR	22-JUL-1999;	99US-0145192P.	PR	21-OCT-1999;	99US-0160815P.
PR	23-JUL-1999;	99US-0145145P.	PR	22-OCT-1999;	99US-0160980P.
PR	23-JUL-1999;	99US-0145218P.	PR	22-OCT-1999;	99US-0160981P.
PR	23-JUL-1999;	99US-0145224P.	PR	22-OCT-1999;	99US-0160989P.
PR	26-JUL-1999;	99US-0145276P.	PR	25-OCT-1999;	99US-0161404P.
PR	27-JUL-1999;	99US-0145913P.	PR	25-OCT-1999;	99US-0161405P.
PR	27-JUL-1999;	99US-0145918P.	PR	25-OCT-1999;	99US-0161406P.
PR	27-JUL-1999;	99US-0145919P.	PR	25-OCT-1999;	99US-0161359P.
PR	28-JUL-1999;	99US-0145951P.	PR	26-OCT-1999;	99US-0161360P.
PR	02-AUG-1999;	99US-0146386P.	PR	26-OCT-1999;	99US-0161361P.
PR	02-AUG-1999;	99US-0146388P.	PR	26-OCT-1999;	99US-0161361P.
PR	02-AUG-1999;	99US-0146389P.	PR	28-OCT-1999;	99US-0161920P.
PR	03-AUG-1999;	99US-0147038P.	PR	28-OCT-1999;	99US-0161992P.
PR	04-AUG-1999;	99US-0147204P.	PR	28-OCT-1999;	99US-0161993P.
PR	04-AUG-1999;	99US-0147302P.	PR	29-OCT-1999;	99US-0162142P.

Mon Mar 8 10:13:27 2004

us-10-021-657-2.rag

Query Match		31.6%; Score 971.5; DB 3; Length 537;
Best Local Similarity		40.9%; Pred. NO. 3.3e-89;
Matches 217; Conservative		87; Mismatches 182; Indels 45; Gaps 12;
QY	27	IALLVVLVQSLKQKQPSWPGVIGATVEOLRNVRHMDLVGYL---SRHRT 82
DB	7	IMILSAITAYFLMLTISICLKGPRVWPIIGSLPGLIENCERHMDWISNLRACSTYGT 66
QY	83	VTVDMEFTS-----YTIADPVNVVHVLKTNFTNYPKGIYVRSYMDVLLGDGIFNADGEL 137
DB	67	CICAIPLAKKQGLVTVTCDPRLNEHLKNRFDNYPKGPWTQAVFHDLLGGQIFNSDGD 126
QY	138	WRKQRTASFEFASKNLRDPSA-IVPREYSLKUSGILSQASKAGKVVDQMQLYMRWILDS 196
DB	127	WLFQRTAALEFTRTLQOAMARWNRRAIKRLPLFLENARLGSBPDLDQLLRLLRTFDN 186
QY	197	ICKVGVGEIGTSLDPENSPQAQDAANIITLRFIDP--LWRIKRFPHVGEALLAQ 254
DB	187	ICGLTFGKDPRTCAPGLPVNTFAVDRAEASLQRFILPEILWKFKRWLRGLVSLTR 246
QY	255	SIKLVDFTYSVIRRRKAEIVEVRASGQKQKMDHILSRFIELGEAGDDGGGFGDDKSLR 314
DB	247	SLVQVDNYLSEIITRKEENMTQHNNKHH---DGLLSREIKKES-----YSDET 296
QY	315	DVVLNFIAGRDTAITLSWFTWAMSHPDVAEKLRLBELCAFEARAREEGVTILVLCGGA 374
DB	297	RVALNFTLAGRDTSSVALSWFFWLITQHPAIEDKILREICTVLVE-TRGDDVALW----- 350
QY	375	DADDKFAARVAQFAGLLTVDSLGKLVYLHACVTETLRLYPAPQDPKGILEDVLPDGT 434
DB	351	-TDEP-----LSCEELDRIVFLKAAISETLRLYPSVPEDSKRAVKDDVLPDGT 397
QY	435	KVRAGGMVTVPYSGMRMEYNWGEDAASFRPERWINE--GAPENASPPKFTAFQAGPRI 492
DB	398	FVPAGSITISYSGRMKSTWGEDCLEFPERWISQDGRFINHDPFKFVAFNAGPRI 457
QY	493	CLKQDSAYLQWK-MALAILFRFYSFRLLGHPVQYRMWTLISMAHGKLVKV 542
DB	458	CLGKDLAYLQWKSIASAVLLR-HRLTVTGKHEQKVSLLTFMKYGLLVNV 507
RESULT 5		
AAG36728		ID AAG36728 standard; protein; 530 AA.
XX	AC	AAG36728;
XX	AC	AAG36728;
XX	AC	AAG36728;
DT	18-OCT-2000	(first entry)
XX	XX	Arabidopsis thaliana protein fragment SEQ ID NO: 45051.
DE	DE	Protein identification; signal transduction pathway; metabolic pathway;
KW	KW	hybridisation assay; genetic mapping; gene expression control; promoter;
KW	KW	termination sequence.
XX	XX	Arabidopsis thaliana.
OS	OS	EP1033405-A2.
PN	PN	06-SEP-2000.
PD	PD	25-FEB-2000; 2000EP-00301439.
XX	XX	25-FEB-1999; 99US-0121825P.
PR	PR	05-MAR-1999; 99US-0123180P.
PR	PR	09-MAR-1999; 99US-0123548P.
PR	PR	23-MAR-1999; 99US-0125788P.
PR	PR	25-MAR-1999; 99US-0126264P.
PR	PR	29-MAR-1999; 99US-0126785P.
PR	PR	01-APR-1999; 99US-0127462P.
PR	PR	06-APR-1999; 99US-0128234P.
PR	PR	08-APR-1999; 99US-0128714P.
PR	PR	16-APR-1999; 99US-0129845P.

19-APR-1999; 99US-0130077P.  
21-APR-1999; 99US-0130449P.  
23-APR-1999; 99US-0130510P.  
23-APR-1999; 99US-0130891P.  
28-APR-1999; 99US-0131449P.  
30-APR-1999; 99US-0132048P.  
30-APR-1999; 99US-0132407P.  
04-MAY-1999; 99US-0132484P.  
05-MAY-1999; 99US-0132485P.  
06-MAY-1999; 99US-0132486P.  
06-MAY-1999; 99US-0132487P.  
07-MAY-1999; 99US-0132863P.  
11-MAY-1999; 99US-0134256P.  
14-MAY-1999; 99US-0134218P.  
14-MAY-1999; 99US-0134219P.  
14-MAY-1999; 99US-0134221P.  
14-MAY-1999; 99US-0134370P.  
18-MAY-1999; 99US-0134768P.  
19-MAY-1999; 99US-0134941P.  
20-MAY-1999; 99US-0135124P.  
21-MAY-1999; 99US-0135353P.  
24-MAY-1999; 99US-0135629P.  
25-MAY-1999; 99US-0136021P.  
27-MAY-1999; 99US-0136392P.  
28-MAY-1999; 99US-0136782P.  
01-JUN-1999; 99US-0137222P.  
03-JUN-1999; 99US-0137528P.  
04-JUN-1999; 99US-0137502P.  
07-JUN-1999; 99US-0137724P.  
08-JUN-1999; 99US-0138094P.  
10-JUN-1999; 99US-0138540P.  
10-JUN-1999; 99US-0138847P.  
14-JUN-1999; 99US-0139119P.  
16-JUN-1999; 99US-0139452P.  
16-JUN-1999; 99US-0139453P.  
17-JUN-1999; 99US-0139492P.  
18-JUN-1999; 99US-0139454P.  
18-JUN-1999; 99US-0139455P.  
18-JUN-1999; 99US-0139456P.  
18-JUN-1999; 99US-0139457P.  
18-JUN-1999; 99US-0139458P.  
18-JUN-1999; 99US-0139459P.  
18-JUN-1999; 99US-0139460P.  
18-JUN-1999; 99US-0139461P.  
18-JUN-1999; 99US-0139462P.  
18-JUN-1999; 99US-0139463P.  
18-JUN-1999; 99US-0139750P.  
18-JUN-1999; 99US-0139763P.  
21-JUN-1999; 99US-0139817P.  
22-JUN-1999; 99US-0139899P.  
23-JUN-1999; 99US-0140353P.  
23-JUN-1999; 99US-0140354P.  
24-JUN-1999; 99US-0140695P.  
28-JUN-1999; 99US-0140823P.  
29-JUN-1999; 99US-0140991P.  
30-JUN-1999; 99US-0141287P.  
01-JUL-1999; 99US-0141842P.  
01-JUL-1999; 99US-0142154P.  
02-JUL-1999; 99US-0142055P.  
06-JUL-1999; 99US-0142390P.  
08-JUL-1999; 99US-0142803P.  
09-JUL-1999; 99US-0142920P.  
12-JUL-1999; 99US-0142977P.  
13-JUL-1999; 99US-0143542P.  
14-JUL-1999; 99US-0143624P.  
15-JUL-1999; 99US-0144005P.  
16-JUL-1999; 99US-0144085P.  
16-JUL-1999; 99US-0144086P.  
19-JUL-1999; 99US-0144325P.  
19-JUL-1999; 99US-0144331P.  
19-JUL-1999; 99US-0144332P.  
19-JUL-1999; 99US-0144333P.  
19-JUL-1999; 99US-0144334P.

PR	19-JUL-1999;	99US-0144335P.	PR	14-OCT-1999;	99US-0159330P.
PR	20-JUL-1999;	99US-0144352P.	PR	14-OCT-1999;	99US-0159331P.
PR	20-JUL-1999;	99US-0144632P.	PR	14-OCT-1999;	99US-0159637P.
PR	20-JUL-1999;	99US-0144884P.	PR	14-OCT-1999;	99US-0159638P.
PR	21-JUL-1999;	99US-0144814P.	PR	18-OCT-1999;	99US-0159584P.
PR	21-JUL-1999;	99US-0145086P.	PR	21-OCT-1999;	99US-0160741P.
PR	21-JUL-1999;	99US-0145088P.	PR	21-OCT-1999;	99US-0160767P.
PR	22-JUL-1999;	99US-0145085P.	PR	21-OCT-1999;	99US-0160768P.
PR	22-JUL-1999;	99US-0145087P.	PR	21-OCT-1999;	99US-0160770P.
PR	22-JUL-1999;	99US-0145089P.	PR	21-OCT-1999;	99US-0160814P.
PR	22-JUL-1999;	99US-0145192P.	PR	21-OCT-1999;	99US-0160815P.
PR	23-JUL-1999;	99US-0145145P.	PR	22-OCT-1999;	99US-0160980P.
PR	23-JUL-1999;	99US-0145218P.	PR	22-OCT-1999;	99US-0160981P.
PR	23-JUL-1999;	99US-0145224P.	PR	22-OCT-1999;	99US-0160989P.
PR	26-JUL-1999;	99US-0145276P.	PR	25-OCT-1999;	99US-0161404P.
PR	27-JUL-1999;	99US-0145313P.	PR	25-OCT-1999;	99US-0161405P.
PR	27-JUL-1999;	99US-0145318P.	PR	25-OCT-1999;	99US-0161406P.
PR	27-JUL-1999;	99US-0145319P.	PR	26-OCT-1999;	99US-0161359P.
PR	28-JUL-1999;	99US-0145351P.	PR	26-OCT-1999;	99US-0161360P.
PR	02-AUG-1999;	99US-0146386P.	PR	26-OCT-1999;	99US-0161361P.
PR	02-AUG-1999;	99US-0146388P.	PR	26-OCT-1999;	99US-0161362P.
PR	02-AUG-1999;	99US-0146389P.	PR	28-OCT-1999;	99US-0161992P.
PR	03-AUG-1999;	99US-0147038P.	PR	28-OCT-1999;	99US-0161993P.
PR	04-AUG-1999;	99US-0147204P.	PR	29-OCT-1999;	99US-0162142P.
PR	04-AUG-1999;	99US-0147302P.			
PR	05-AUG-1999;	99US-0147192P.			
PR	05-AUG-1999;	99US-0147260P.			
PR	06-AUG-1999;	99US-0147303P.			
PR	06-AUG-1999;	99US-0147416P.			
PR	08-AUG-1999;	99US-0147493P.			
PR	09-AUG-1999;	99US-0147935P.			
PR	10-AUG-1999;	99US-0148171P.			
PR	11-AUG-1999;	99US-0148319P.			
PR	12-AUG-1999;	99US-0148341P.			
PR	13-AUG-1999;	99US-0148565P.			
PR	13-AUG-1999;	99US-0148684P.			
PR	16-AUG-1999;	99US-0149368P.			
PR	17-AUG-1999;	99US-0149317P.			
PR	18-AUG-1999;	99US-0149426P.			
PR	20-AUG-1999;	99US-0149722P.			
PR	20-AUG-1999;	99US-0149723P.			
PR	20-AUG-1999;	99US-0149929P.			
PR	23-AUG-1999;	99US-0149902P.			
PR	23-AUG-1999;	99US-0149930P.			
PR	25-AUG-1999;	99US-0150566P.			
PR	26-AUG-1999;	99US-0150884P.			
PR	27-AUG-1999;	99US-0151065P.			
PR	27-AUG-1999;	99US-0151066P.			
PR	27-AUG-1999;	99US-0151080P.			
PR	30-AUG-1999;	99US-0151303P.			
PR	31-AUG-1999;	99US-0151438P.			
PR	01-SEP-1999;	99US-0151930P.			
PR	07-SEP-1999;	99US-0152363P.			
PR	10-SEP-1999;	99US-0153070P.			
PR	13-SEP-1999;	99US-0153758P.			
PR	15-SEP-1999;	99US-0154018P.			
PR	16-SEP-1999;	99US-0154039P.			
PR	20-SEP-1999;	99US-0154779P.			
PR	22-SEP-1999;	99US-0155139P.			
PR	23-SEP-1999;	99US-0155486P.			
PR	24-SEP-1999;	99US-0155659P.			
PR	28-SEP-1999;	99US-0156458P.			
PR	29-SEP-1999;	99US-0156596P.			
PR	04-OCT-1999;	99US-0157117P.			
PR	05-OCT-1999;	99US-0157533P.			
PR	06-OCT-1999;	99US-0157865P.			
PR	07-OCT-1999;	99US-0158029P.			
PR	08-OCT-1999;	99US-0158232P.			
PR	12-OCT-1999;	99US-0158369P.			
PR	13-OCT-1999;	99US-0159293P.			
PR	13-OCT-1999;	99US-0159294P.			
PR	13-OCT-1999;	99US-0159295P.			
PR	14-OCT-1999;	99US-0159323P.			
<p>Query Match 31.5%; Score 970.5; DB 3; Length 530;            Best Local Similarity 41.0%; Pred. No. 4e-89;            Matches 217; Conservative 86; Mismatches 181; Indels 45; Gaps 12;</p>					
QY	29	LLLVLSLWLVQSLRKQKSPRWFVIGATVEQLNRYHRMHDWLVGYL-----SHRVTVT 84			
DB	2	IILSAITAYFLWLTFTSRCLKGPVWPILGSLPLGLIENCERMHDWISDNLRACSGYTQTCI 61			
QY	85	VDMPTFS-----YTYIADPVNTEHVILKTNFTNYPKGVIVRSYMDVLVLGDIENAGELWR 139			
DB	62	CAIPELAKKQGLVTVTCDDPRNLEHLKKNRFDNYPKGPWTQAVFHLGGLGGINSGDWTWL 121			
QY	140	KOKTASFEFASKNLRFDSA-IVFREYSILGSLISQASKAGKGVMDQELYNRMVTLDSIC 198			
DB	122	FORKTAALFTRTLRQAVARNWNRKILRFLPILENARLGSEPIDLQDLALLRLTFDNC 181			
QY	199	KVGFGEIGTSLDLPENSFAQAFDAANIITLRFIDP--LWRIKFFHVGEALLAQST 256			
DB	182	GLTFGKDPRTCAPLPVNTFAVAFDEATASLQRFILPEILWKFKEWLRLGLEVSLTRSL 241			
QY	257	KLVDFTYSVIRRRKAEIIVEVRASQKQKMKHDIILSRFIELGHAGDDGGFGDDKSLRDV 316			
DB	242	VQVDNYLSBIITTRKEENMTQHNNGKH---DOLLSRFICKKES-----YSEDTLQRV 291			
QY	317	VLFNVIAGRTTATTLISWFTHMAMSHPDVAEKLRLRELCAPAEARAREEGVTLVLCGGADA 376			
DB	292	ALNFIAGRTSSVALSWFFWLLITQHPAIEDKILRICIVLVE-TRGDDVALW-----T 344			
QY	377	DDKAFARVAQAQAGLITYDSLGLKLYLHACVTETLRLYPVAPQDPKGILEDVLPDGTKV 436			
DB	345	DEP-----LSCELDRLVFLKAALSETLRLYSPVSPEDSKRAVKDDVLPDGTFFV 392			
QY	437	RAGMVTYYPYSGRMEYKWPDAASFRRERWNEED--GAFRNASPKFTAFQAGPRICL 494			
DB	393	PAGSSITYSYISAGRMKSTWGEDCLFEPKPERWISQSDGGRFINHDPFKFVAFNAGPRICL 452			
QY	495	GKDSAYLQMK-MALAILFRYSPRLLEGHVPQVYRMMTILSMAHGLKVRV 542			
DB	453	GKDLAYLQMKSIASAVLLR-HRLTVTVTGKVEQKNSLTFMKYGLLVNV 500			
<p>RESULT 6            AAG54203            ID AAG54203 standard; protein; 523 AA.            XX AAG54203;            AC AAG54203;            XX AAG54203;            DT 18-OCT-2000 (first entry)</p>					



PR	27-AUG-1999;	99US-0151080P.	Query Match	31.5%; Score 969.5; DB 3; Length 523;
PR	30-AUG-1999;	99US-0151303P.	Best Local Similarity	39.7%; Pred. No. 5e-89;
PR	31-AUG-1999;	99US-0151438P.	Matches 211; Conservative	99; Mismatches 173; Indels 49; Gaps 12;
PR	01-SEP-1999;	99US-0151930P.		
PR	07-SEP-1999;	99US-0152363P.		
PR	10-SEP-1999;	99US-0153070P.		
PR	13-SEP-1999;	99US-0153758P.		
PR	15-SEP-1999;	99US-0154018P.		
PR	16-SEP-1999;	99US-0154039P.		
PR	20-SEP-1999;	99US-0154779P.		
PR	22-SEP-1999;	99US-0155139P.		
PR	23-SEP-1999;	99US-0155486P.		
PR	24-SEP-1999;	99US-0155659P.		
PR	28-SEP-1999;	99US-0156458P.		
PR	29-SEP-1999;	99US-0156596P.		
PR	04-OCT-1999;	99US-0157117P.		
PR	05-OCT-1999;	99US-015753P.		
PR	06-OCT-1999;	99US-0157865P.		
PR	07-OCT-1999;	99US-0158029P.		
PR	08-OCT-1999;	99US-0158232P.		
PR	12-OCT-1999;	99US-0158369P.		
PR	13-OCT-1999;	99US-0159293P.		
PR	13-OCT-1999;	99US-0159294P.		
PR	13-OCT-1999;	99US-0159295P.		
PR	14-OCT-1999;	99US-0159329P.		
PR	14-OCT-1999;	99US-0159330P.		
PR	14-OCT-1999;	99US-0159331P.		
PR	14-OCT-1999;	99US-0159637P.		
PR	14-OCT-1999;	99US-0159638P.		
PR	18-OCT-1999;	99US-0159584P.		
PR	21-OCT-1999;	99US-0160741P.		
PR	21-OCT-1999;	99US-0160767P.		
PR	21-OCT-1999;	99US-0160768P.		
PR	21-OCT-1999;	99US-0160770P.		
PR	21-OCT-1999;	99US-0160814P.		
PR	21-OCT-1999;	99US-0160815P.		
PR	22-OCT-1999;	99US-0160980P.		
PR	22-OCT-1999;	99US-0160981P.		
PR	22-OCT-1999;	99US-0160989P.		
PR	25-OCT-1999;	99US-0161404P.		
PR	25-OCT-1999;	99US-0161405P.		
PR	25-OCT-1999;	99US-0161406P.		
PR	26-OCT-1999;	99US-0161359P.		
PR	26-OCT-1999;	99US-0161360P.		
PR	26-OCT-1999;	99US-0161361P.		
PR	28-OCT-1999;	99US-0161520P.		
PR	28-OCT-1999;	99US-0161992P.		
PR	28-OCT-1999;	99US-0161993P.		
PR	29-OCT-1999;	99US-0162142P.		

QY	27	IALLVLSLWVQSLRQKQPRSPVIGATVEQLNRYHRMHDVLVGL----	SRHRT	82
Db	10	LTIVTVIIVFSLR-----RYSKGRVWPVIGSLPALITNAHRMHDFIADNRCGGTYQT	66	
QY	83	VTVDMPFTS-----YTYIADPVNVEHVLKTNFTNPKGIYVRSYMDVLGDIIPNADGEL	137	
Db	67	CIPFIPLAKQGHVTVTCDPKLEHLKTRFDNPKYKPSQSVFHDLLGDGIFNSDGT	126	
QY	138	WRKORKTASFEFASKNLRDFA-IVFREYSLKLSGILSQASKAKVVDQMQLYMRVTLDS	196	
Db	127	WRFQKTAALFEFTRTLQAMRWVDRAIKNRLVPILESARSRAEPIDLQVLLRLTFDN	186	
QY	197	ICKVGFGEIGTSPDLSPNSFACAFDAANIIILREIDP--LWRIKRFHFVGVSEALLAQ	254	
Db	187	ICGLTFQKDKPRTLSPEFPENGFAVDFGATFATLQRFIMEFIIWKIKWILGLGDDMSR	246	
QY	255	SIKLVDFTYSVIRRRKAEIIVEVRASGKQKMKH-DILSRFIELGEAGDDGGGDDKSL	313	
Db	247	SISHVDNLSFIINTRKLELL---GQQQDSRHDDLRLSRPMKKES-----YSDKYL	295	

QY	314	ROYVNFVIAGRDITATTLSWFTHMAMSHDVAEKLRRELCAFEAEARARBEQVTLVLCGG	373	
Db	296	KYVALNFILAGRDITSSVAMSWFFWLVSLNPRVEKIINEIC-----TILI---	340	
QY	374	ADADDKAFARVAQFAG-LTYDSLGLKLVYLHACVTETLRLYPADPPKGILEDVLPD	432	
Db	341	-----KTRDINVSQWTDPLTDFEIDQLVYLKAALSETLRLYSPEDSKFVVANDVLPD	395	
QY	433	GTKVRAGMVTYVPYSNGRMEYNWGPDAASFRPERMINEDGAFRNASFEKFTAFQAGPRI	492	
Db	396	GTFVPGSNVTYSTYSVGRMKFIWGEDCLSEKPERWL-EESRDEKCMQYKFVAENAGPRI	454	
QY	493	CLGRDSAYLOWKMALAILFRFYSRLLLEGHPVOYRMMTILSMHAGLKVRSR	544	
Db	455	CLGHDLAYLQMKSIATASILLRHLITVAPGRHVEQKMSLTLFMKFGKMDVHK	506	

RESULT 7

AAG24684

ID AAG24684 standard; protein; 522 AA.

XX

AC AAG24684;

XX

DT 17-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 28454.

XX

KW Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.

XX

OS Arabidopsis thaliana.

XX

PN EP1033405-A2.

XX

PD 06-SEP-2000.

XX

PF 25-FEB-2000; 2000EP-00301439.

XX

PR 25-FEB-1999; 99US-0121825P.

PR

PR 05-MAR-1999; 99US-0123180P.

PR

PR 09-MAR-1999; 99US-0123548P.

PR

PR 23-MAR-1999; 99US-0125788P.

PR

PR 25-MAR-1999; 99US-0126264P.

PR

PR 29-MAR-1999; 99US-0126785P.

PR

PR 01-APR-1999; 99US-0127462P.

PR

PR 06-APR-1999; 99US-0128234P.

PR

PR 08-APR-1999; 99US-0128714P.

PR

PR 16-APR-1999; 99US-0129845P.

PR

PR 19-APR-1999; 99US-0130077P.

PR

PR 21-APR-1999; 99US-0130449P.

PR

PR 23-APR-1999; 99US-0130510P.

PR

PR 23-APR-1999; 99US-0130891P.

PR

PR 28-APR-1999; 99US-0134449P.

PR

PR 30-APR-1999; 99US-0132048P.

PR

PR 30-APR-1999; 99US-0132407P.

PR

PR 04-MAY-1999; 99US-0132484P.

PR

PR 05-MAY-1999; 99US-0132485P.

PR

PR 06-MAY-1999; 99US-0132486P.

PR

PR 07-MAY-1999; 99US-0132487P.

PR

PR 11-MAY-1999; 99US-0132863P.

PR

PR 14-MAY-1999; 99US-0134218P.

PR

PR 14-MAY-1999; 99US-0134219P.

PR

PR 14-MAY-1999; 99US-0134221P.

PR

PR 14-MAY-1999; 99US-0134370P.

PR

PR 14-MAY-1999; 99US-0134376P.

PR

PR 19-MAY-1999; 99US-0134941P.

PR

PR 20-MAY-1999; 99US-0135124P.

PR

PR 21-MAY-1999; 99US-0135353P.

PR

PR 24-MAY-1999; 99US-0135629P.

PR

PR 25-MAY-1999; 99US-0136021P.

PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142058P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144003P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144637P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 13-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 01-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-015753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159283P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160800P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161992P.  
PR 28-OCT-1999; 99US-0161993P.

```
PR 29-OCT-1999; 99US-0162142P.
Query Match 31.2%; Score 961.5; DB 3; Length 522;
Best Local Similarity 39.4%; Pred. No. 3.2e-88;
Matches 209; Conservative 98; Mismatches 177; Indels 47; Gaps 11;
QY 27 IALLVILSWILVQWLSRKKQKPRSWPVGIVTVEQLRNYHMHMLVGYL-----SRHRT 82
DB 9 LLIIVYILWFVSLR---RSYKPRVWPLVGLPALITNAHRMHDFIADNLNRCGGTYQT 65
QY 83 VTVDMPFTS-----YTIADPVNVEHVLTNTNYPKGIYVESYMDVLLGDGFIENADGEL 137
DB 66 CIEPIPLAKQGHVTVTCDPKNLEHLKTRFDNVPKGSWSQSVFHDLLDGDGFIENSDGT 125
QY 138 WKQKRTAGFEASKNLRFDSA-IVPREYSLKSLGILSQASKAGKVVDVMDQELMYMTLDS 196
DB 126 WFOKTALETTTRILQAMARWVDBAKNRLVPILESARSAEPIIDQDVLRLITFDN 185
QY 197 ICKVGFGEIGTSPDLPPNSFAQFADAANIITLRFIDP--LWRIKRFHVHVSALLAQ 254
DB 186 ICGLTFGKDPRLSPPEFFENGFAVAFDGATEATLQRFIMPEFIWKIRKWLRLGLEDDMR 245
QY 255 SIKLVDEFYVTRRKARIVEVRASQKQKMKHDILSRFIELGAGDDGGFGDDKSLR 314
DB 246 SISHVDNLYSEIINTRKELLGQQQDGRH---DCLLSRFMKKES-----YSDKYLK 295
QY 315 DVVLFNFIAGROTTATTLGFWTHMAMSHPDVAEKRLRELCAFEARAREEGVTLVLCGGA 374
DB 296 YVALNFILAGNTSSVAMSWFFLWVLSINPRVEKIINEIC-----TILI--- 339
QY 375 DADKKAFAARVAQFAG-LTYDSLGLKLYVHACVTTLRLYPVAVPQPKGILLEDDVLPDG 433
DB 340 ----KTRDTNVSKWTDEPLTFEIDQLVYLKAALSETLRLYSPEDSKFVVANDVLPDG 395
QY 434 TKVRAGGWTVYPYSGRMENYNGPDAASPRPERWINEDEGAPRNAGPFKFTAFQAGPRIC 493
DB 396 TFPVSGSNVTYISYVGRMKFIWGEDCLEFKPERWL-EESRDEKQKQYKEVAFNAGPRIC 454
QY 494 LGKDSAYLQMKALAILFRYSPRLLEGHVPVQVYMMTILSMAGHLKVVRVS 544
DB 455 LGKDLAYLQMKSTITALLRHRLTVAPGRHVEQKMSLTLPWKFGKMDVHK 505
RESULT 8
AAG24683
ID AAG24683 standard; protein; 527 AA.
XX AC AAG24683;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 28453.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-00301439.
XX QY 25-FEB-1999; 99US-0121825P.
XX PR 05-MAR-1999; 99US-0123180P.
XX PR 23-MAR-1999; 99US-0123548P.
XX PR 29-MAR-1999; 99US-0125788P.
XX PR 29-MAR-1999; 99US-0126264P.
XX PR 01-APR-1999; 99US-0126785P.
XX PR 06-APR-1999; 99US-0127462P.
XX PR 06-APR-1999; 99US-0128234P.
99US-0128714P.
99US-0129845P.
99US-0130077P.
99US-0130449P.
99US-0130510P.
99US-0130891P.
99US-0131449P.
99US-0132048P.
99US-0132407P.
99US-0132484P.
99US-0132485P.
99US-0132486P.
99US-0132487P.
99US-0132863P.
99US-0134256P.
99US-0134258P.
99US-0134219P.
99US-0134221P.
99US-0134370P.
99US-0134768P.
99US-0134941P.
99US-0135124P.
99US-0135353P.
99US-0135629P.
99US-0136021P.
99US-0136392P.
99US-0136782P.
99US-0137222P.
99US-0137528P.
99US-0137503P.
99US-0137724P.
99US-0138094P.
99US-0138540P.
99US-0138847P.
99US-0139119P.
99US-0139452P.
99US-0139453P.
99US-0139454P.
99US-0139455P.
99US-0139456P.
99US-0139457P.
99US-0139458P.
99US-0139459P.
99US-0139460P.
99US-0139461P.
99US-0139462P.
99US-0139463P.
99US-0139750P.
99US-0139763P.
99US-0139817P.
99US-0139899P.
99US-0140353P.
99US-0140354P.
99US-0140695P.
99US-0140823P.
99US-0140981P.
99US-0141287P.
99US-0141842P.
99US-0142154P.
99US-0142055P.
99US-0142390P.
99US-0142803P.
99US-0142920P.
99US-0142977P.
99US-0143542P.
99US-0143624P.
99US-0144005P.
99US-0144085P.
99US-0144866P.
99US-0144325P.
99US-0144331P.
99US-0144332P.
```

RESULT 9  
AAG36729  
ID AAG  
XX  
AC AAG

XX 18-OCT-2000 (first entry)  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 45052.  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KW Hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX Arabidopsis thaliana.  
XX EP1033405-A2.  
XX 06-SEP-2000.  
XX 25-FEB-2000; 2000EP-00301439.  
XX 25-FEB-1999; 99US-0121825P.  
XX 05-MAR-1999; 99US-0123180P.  
XX 09-MAR-1999; 99US-0123548P.  
XX 23-MAR-1999; 99US-0125788P.  
XX 25-MAR-1999; 99US-0126264P.  
XX 29-MAR-1999; 99US-0126785P.  
XX 01-APR-1999; 99US-0127462P.  
XX 06-APR-1999; 99US-0128234P.  
XX 08-APR-1999; 99US-0128714P.  
XX 16-APR-1999; 99US-0129845P.  
XX 19-APR-1999; 99US-0130077P.  
XX 21-APR-1999; 99US-0130449P.  
XX 23-APR-1999; 99US-0130510P.  
XX 28-APR-1999; 99US-0130891P.  
XX 30-APR-1999; 99US-0131449P.  
XX 30-APR-1999; 99US-0132048P.  
XX 04-MAY-1999; 99US-0132407P.  
XX 05-MAY-1999; 99US-0132485P.  
XX 06-MAY-1999; 99US-0132486P.  
XX 06-MAY-1999; 99US-0132487P.  
XX 07-MAY-1999; 99US-0132863P.  
XX 11-MAY-1999; 99US-0134256P.  
XX 14-MAY-1999; 99US-0134218P.  
XX 14-MAY-1999; 99US-0134219P.  
XX 14-MAY-1999; 99US-0134221P.  
XX 14-MAY-1999; 99US-0134370P.  
XX 18-MAY-1999; 99US-0134768P.  
XX 19-MAY-1999; 99US-0134941P.  
XX 20-MAY-1999; 99US-0135124P.  
XX 21-MAY-1999; 99US-0135353P.  
XX 24-MAY-1999; 99US-0135629P.  
XX 25-MAY-1999; 99US-0136021P.  
XX 27-MAY-1999; 99US-0136392P.  
XX 28-MAY-1999; 99US-0136782P.  
XX 01-JUN-1999; 99US-0137222P.  
XX 03-JUN-1999; 99US-0137528P.  
XX 04-JUN-1999; 99US-0137502P.  
XX 07-JUN-1999; 99US-0137724P.  
XX 08-JUN-1999; 99US-0138094P.  
XX 10-JUN-1999; 99US-0138540P.  
XX 10-JUN-1999; 99US-0138847P.  
XX 14-JUN-1999; 99US-0139119P.  
XX 16-JUN-1999; 99US-0139452P.  
XX 16-JUN-1999; 99US-0139453P.  
XX 17-JUN-1999; 99US-0139492P.  
XX 18-JUN-1999; 99US-0139454P.  
XX 18-JUN-1999; 99US-0139455P.  
XX 18-JUN-1999; 99US-0139456P.  
XX 18-JUN-1999; 99US-0139457P.  
XX 18-JUN-1999; 99US-0139458P.  
XX 18-JUN-1999; 99US-0139459P.  
XX 18-JUN-1999; 99US-0139460P.  
XX 18-JUN-1999; 99US-0139461P.  
XX 18-JUN-1999; 99US-0139462P.  
XX 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.

PR	27-AUG-1999;	99US-0151065P.	Db	238	KKBS-----YSDTLQVALNFILAGDITSSVALSWFFWLLITQHPAIEDKILREICTV	290
PR	27-AUG-1999;	99US-0151066P.				
PR	27-AUG-1999;	99US-0151080P.	QY	357	EAERAREEGVTLVLCGGADADKAFARVAQFAGLLTYDSLKIVYLHACVTETLRLYPA	416
PR	30-AUG-1999;	99US-01511303P.				
PR	31-AUG-1999;	99US-0151438P.	Db	291	LVE-TRGDDVALW-----TDEP-----LSCELDRLVFLKAALSETLRLYPS	331
PR	01-SEP-1999;	99US-0151930P.				
PR	07-SEP-1999;	99US-01522363P.	QY	417	VPQDPKGILEDVLPDGTIKVRAGMVTVVPYSGMRMEYNWGPDAASFPRPERWINE--GA	474
PR	10-SEP-1999;	99US-0153070P.				
PR	13-SEP-1999;	99US-0153758P.				
PR	15-SEP-1999;	99US-0154018P.	Db	332	VEEDSKRAVKODVLPDGTFPVAGSSITTSIYAGRMKSTWGEDCIEFKPERMISQSDGGR	391
PR	16-SEP-1999;	99US-0154039P.				
PR	20-SEP-1999;	99US-0154779P.				
PR	22-SEP-1999;	99US-0155139P.	QY	475	FNASPFKETAFAQGPRIKLGKDSAYLQMK-MALAILFRYSFRLLGHPVQVRRMTIIS	533
PR	23-SEP-1999;	99US-0155438P.				
PR	24-SEP-1999;	99US-0155659P.	Db	392	FINHDPFKFVANAGPRIKLGKDLAYLQMKSIASAVLLR-HRLTVVTGHHKVEQKMSLTIF	450
PR	28-SEP-1999;	99US-0156458P.				
PR	29-SEP-1999;	99US-0156596P.	QY	534	MAHGLKRVV	542
PR	04-OCT-1999;	99US-0157117P.	Db	451	MKYGLLVNV	459
PR	05-OCT-1999;	99US-0157753P.				
PR	06-OCT-1999;	99US-0157865P.				
PR	07-OCT-1999;	99US-0158029P.				
PR	08-OCT-1999;	99US-0158232P.				
PR	12-OCT-1999;	99US-0158369P.				
PR	13-OCT-1999;	99US-0159293P.				
PR	13-OCT-1999;	99US-0159294P.				
PR	13-OCT-1999;	99US-0159329P.				
PR	14-OCT-1999;	99US-0159330P.				
PR	14-OCT-1999;	99US-0159331P.				
PR	14-OCT-1999;	99US-0159637P.				
PR	14-OCT-1999;	99US-0159638P.				
PR	18-OCT-1999;	99US-0159584P.				
PR	21-OCT-1999;	99US-0160741P.				
PR	21-OCT-1999;	99US-0160767P.				
PR	21-OCT-1999;	99US-0160768P.				
PR	21-OCT-1999;	99US-0160770P.				
PR	21-OCT-1999;	99US-0160814P.				
PR	21-OCT-1999;	99US-0160815P.				
PR	22-OCT-1999;	99US-0160980P.				
PR	22-OCT-1999;	99US-0160981P.				
PR	22-OCT-1999;	99US-0160989P.				
PR	25-OCT-1999;	99US-0161404P.				
PR	25-OCT-1999;	99US-0161405P.				
PR	25-OCT-1999;	99US-0161406P.				
PR	26-OCT-1999;	99US-0161359P.				
PR	26-OCT-1999;	99US-0161360P.				
PR	26-OCT-1999;	99US-0161361P.				
PR	28-OCT-1999;	99US-0161920P.				
PR	28-OCT-1999;	99US-0161992P.				
PR	28-OCT-1999;	99US-0161993P.				
PR	29-OCT-1999;	99US-0162142P.				
Query Match 29.6%; Score 911.5; DB 3; Length 489;						
Best Local Similarity 41.9%; Pred.No. 3.6e-83;						
Matches 205; Conservative 77; Mismatches 162; Indels 45; Gaps 12;						
QY	69	MHDVLGVYL-----SRHRTVTVDMPETS-----YTVIADPVNVHVKTNFTNPKGIYVR	119			
Db	1	MHDWISDNLKACSGYQTCICAIPLAKKQGLVTVTCDPRLNHLKRNFDNPKGPTWQ	60			
QY	120	SYMDVLGGIFNAGELMRKORTASPEFAKKNLRDPSA-IVFREYSKLKSLGILSQASK	178			
Db	61	AVFHDLGGQIFNSDGDWLFORTKAALEFTTTLRQAMRWVNRKILRFLPILNARL	120			
QY	179	AKQVDMQELNRYMTLDSTCKVGFGEIGTLPDLPENSFAQDAANIIILRFTDP--	236			
Db	121	GSEPIDQLQLRLTLDNFCGLTFGKDPRTCAPGLPVNTFAVAFRATASLQRFILPEI	180			
QY	237	LWRIKFFHVGSEALLAQKILVDEFTYSVIRRRKAEIIVEVRASGKQEKMKHDIILSRFTE	296			
Db	181	LWKFKWLRGLGVSTRLVQVDNLYLSIIITRKEMMTQHNKGKH--DDLLSRFIK	237			
QY	297	LGEAGDGGGFGDKSLRDVNLVFIAGRDITATTLISWNTHMAMSHPDVAEKILRELCAF	356			

RESULT 10

AAG54204

ID AAG54204 standard; protein; 475 AA.

XX AC AAG54204;

XX DT 18-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 69088.

XX KW Protein identification; signal transduction pathway; metabolic pathway;  
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX KW termination sequence.

OS Arabidopsis thaliana.

PN EPI033405-A2.

XX FD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-00301439.

XX PR 25-FEB-1999; 99US-0121825P.

XX PR 05-MAR-1999; 99US-0123180P.

XX PR 09-MAR-1999; 99US-0123588P.

XX PR 23-MAR-1999; 99US-0125788P.

XX PR 25-MAR-1999; 99US-0126264P.

XX PR 29-MAR-1999; 99US-0126785P.

XX PR 01-APR-1999; 99US-0127462P.

XX PR 06-APR-1999; 99US-0128234P.

XX PR 08-APR-1999; 99US-0128714P.

XX PR 16-APR-1999; 99US-0129845P.

XX PR 19-APR-1999; 99US-0130077P.

XX PR 21-APR-1999; 99US-0130449P.

XX PR 23-APR-1999; 99US-0130510P.

XX PR 23-APR-1999; 99US-0130891P.

XX PR 28-APR-1999; 99US-0131449P.

XX PR 30-APR-1999; 99US-0132048P.

XX PR 30-APR-1999; 99US-0132407P.

XX PR 04-MAY-1999; 99US-0132484P.

XX PR 05-MAY-1999; 99US-0132485P.

XX PR 06-MAY-1999; 99US-0132486P.

XX PR 06-MAY-1999; 99US-0132487P.

XX PR 07-MAY-1999; 99US-0132863P.

XX PR 11-MAY-1999; 99US-0134256P.

XX PR 14-MAY-1999; 99US-0134218P.

XX PR 14-MAY-1999; 99US-0134219P.

XX PR 14-MAY-1999; 99US-0134221P.

XX PR 14-MAY-1999; 99US-0134370P.

XX PR 14-MAY-1999; 99US-0134768P.

XX PR 18-MAY-1999; 99US-0134941P.

XX PR 19-MAY-1999; 99US-0135124P.

XX PR 20-MAY-1999; 99US-0135353P.

XX PR 21-MAY-1999; 99US-0135353P.

99US-0146382P  
99US-0147038P  
99US-0147204P  
99US-0147302P  
99US-0147392P  
99US-0147603P  
99US-0147603P  
99US-014716P  
99US-0147493P  
99US-0147935P  
99US-0148171P  
99US-0148319P  
99US-0148341P  
99US-0148565P  
99US-0148684P  
99US-0149368P  
99US-0149475P  
99US-0149426P  
99US-0149722P  
99US-0149723P  
99US-0149732P  
99US-0149820P  
99US-0149820P  
99US-0149936P  
99US-0150566P  
99US-0150884P  
99US-0151065P  
99US-0151066P  
99US-0151089P  
99US-0151338P  
99US-0151438P  
99US-0151930P  
99US-0152363P  
99US-0153070P  
99US-0153758P  
99US-0154018P  
99US-0154039P  
99US-0154773P  
99US-0155139P  
99US-0155193P  
99US-0155486P  
99US-0155659P  
99US-0156586P  
99US-0156596P  
99US-0157117P  
99US-0157753P  
99US-0157865P  
99US-0158029P  
99US-0158262P  
99US-0159359P  
99US-0159361P  
99US-0159363P  
99US-0159291P  
99US-0159294P  
99US-0159295P  
99US-0159329P  
99US-0159330P  
99US-0159331P  
99US-0159311P  
99US-0159633P  
99US-0159637P  
99US-0159584P  
99US-0160741P  
99US-0160767P  
99US-0160768P  
99US-0160770P  
99US-0160814P  
99US-0160815P  
99US-0160880P  
99US-0160981P  
99US-0160983P  
99US-0161404P  
99US-0161405P  
99US-0161406P  
99US-0161359P  
99US-0161360P  
99US-0161361P  
99US-0161920P

PR	28-OCT-1999;	99US-0161992P.	PR	01-APR-1999;	99US-0127462P.
PR	28-OCT-1999;	99US-0161993P.	PR	06-APR-1999;	99US-0128234P.
PR	29-OCT-1999;	99US-0162142P.	PR	08-APR-1999;	99US-0128714P.
Query Match 29.0%; Score 894; DB 3; Length 475;					
Best Local Similarity 40.2%; Pred. No. 2e-81;					
Matches 197; Conservative 90; Mismatches 157; Indels 46; Gaps 11;					
QY	69	MHDWLVGYL-----SRHRTVTVDMFFTS-----YTYTADPNVVEHVLKTNFTNPKGIVVR 119	PR	28-APR-1999;	99US-0131449P.
DB	1	MHDFIADNLRMCGGTQCTTFPIFFLAKQGHVTVTCDFKNLEHLIKTFEDNPKGPSWQ 60	PR	30-APR-1999;	99US-0132048P.
QY	120	SYMDVLGGDIGFNADGELMRKQRKTASFEFASKNLDFSA-IVFREYSYKLSGILSQASK 178	PR	30-APR-1999;	99US-0132407P.
DB	61	SVFHDLLGGDIGFNSDGTWRFQKTAALBFETRTLQAMARWDRAIKNELVPILESARS 120	PR	03-APR-1999;	99US-0132484P.
QY	179	AGKVDVMQELYMRTMLDSICKVGGFVEIGTLPDLPENSFAQAFDAANIITLRFIDP-- 236	PR	05-MAY-1999;	99US-0132485P.
DB	121	RAEPIDLQDVLLRLTENNICGLTFGKDPRTLSPFEPENGFAVAFDGAATLQRFIMPF 180	PR	06-MAY-1999;	99US-0132486P.
QY	237	LWRIKFFHVGSEALLAQSILKLVDEITYSVIRPKAEIIVEVRASGQEKMKH-DILSRFI 295	PR	07-MAY-1999;	99US-0132863P.
DB	181	IKWIKRWLRUGLSDDSRSLSHVDNLSEILNTRKLELI-----GQQQDESRRHDDLSRFM 236	PR	11-MAY-1999;	99US-0134256P.
QY	296	ELGEAGDGGGFGDDKSLRDVNLVFNFIAGSDTTATTLSWFTNWMSPHDPVAEKLRELCA 355	PR	14-MAY-1999;	99US-0134218P.
DB	237	KKKES-----YSDKYLKVVALLNFIAGSDTSSVAMSWFFWLVSINPRVEEKIINEIC- 288	PR	14-MAY-1999;	99US-0134219P.
QY	356	FEARAREEGVTLVLCGGADADDKAAFAARVAQFAG-LLIYDSLGKLVYLHACVTEFLRY 414	PR	14-MAY-1999;	99US-0134221P.
DB	289	-----TTLI-----KTRDTNYSKWTDEPLTFDEIDQVLVYKLAULSETRLY 329	PR	14-MAY-1999;	99US-0134370P.
QY	415	PAVPQDPKGILEDVLPDGTQKVRAGGMVTVPVYSMGRMEYNWGPDAASPRPERWINEDGA 474	PR	18-MAY-1999;	99US-0134768P.
DB	330	PSVPEBSKFFVANDVLPDGTFFVSGSNVTYISYVGRMKFIWGEDCLCPKPERWL-EESR 388	PR	19-MAY-1999;	99US-0134941P.
QY	475	FRNASPPKFTAFQAGPRICLGKQSAIYLMQKVALAILFRFYSFRLLEGHPVQYRMWTLISM 534	PR	20-MAY-1999;	99US-0135124P.
DB	389	DEKCNQKYFAFNAFAPRICLGKOLAYLQMKSIITASILLRHLRTVAFCHRVEQKQVUTLFM 448	PR	21-MAY-1999;	99US-0135353P.
QY	535	AHGLKRVSVR 544	PR	24-MAY-1999;	99US-0135629P.
DB	449	KFGLKMDVHK 450	PR	25-MAY-1999;	99US-0136031P.
RESULT 11					
AAG24685					
ID	AAG24685 standard; protein; 475 AA.				
AC	AAG24685;				
XX	17-OCT-2000 (first entry)				
DT	Arabidopsis thaliana protein fragment SEQ ID NO: 28455.				
DE	Protein identification; signal transduction pathway; metabolic pathway;				
KW	hybridisation assay; Genetic mapping; gene expression control; promoter;				
KW	termination sequence.				
XX	Arabidopsis thaliana.				
OS	EP1033405-A2.				
PN	06-SEP-2000.				
PD	25-FEB-2000; 2000EP-00301439.				
PF	25-FEB-1999; 99US-0121825P.				
XX	05-MAR-1999; 99US-0123182P.				
PR	09-MAR-1999; 99US-0123548P.				
PR	23-MAR-1999; 99US-0125788P.				
PR	29-MAR-1999; 99US-0126264P.				
PR	29-MAR-1999; 99US-0126785P.				
PR	99US-0128234P.				
PR	99US-0128714P.				
PR	99US-0129845P.				
PR	99US-0130077P.				
PR	99US-0130449P.				
PR	99US-0130510P.				
PR	99US-0130891P.				

PR 29-N

PR	19-JUL-1999;	99US-0144331P.	PR	13-OCT-1999;	99US-0159293P.
PR	19-JUL-1999;	99US-0144332P.	PR	13-OCT-1999;	99US-0159294P.
PR	19-JUL-1999;	99US-0144333P.	PR	13-OCT-1999;	99US-0159295P.
PR	19-JUL-1999;	99US-0144334P.	PR	14-OCT-1999;	99US-0159329P.
PR	19-JUL-1999;	99US-0144335P.	PR	14-OCT-1999;	99US-0159330P.
PR	20-JUL-1999;	99US-0144336P.	PR	14-OCT-1999;	99US-0159331P.
PR	20-JUL-1999;	99US-0144337P.	PR	14-OCT-1999;	99US-0159637P.
PR	20-JUL-1999;	99US-0144338P.	PR	14-OCT-1999;	99US-0159638P.
PR	20-JUL-1999;	99US-0144884P.	PR	18-OCT-1999;	99US-0159584P.
PR	21-JUL-1999;	99US-0144814P.	PR	21-OCT-1999;	99US-0160741P.
PR	21-JUL-1999;	99US-0145086P.	PR	21-OCT-1999;	99US-0160767P.
PR	21-JUL-1999;	99US-0145088P.	PR	21-OCT-1999;	99US-0160770P.
PR	22-JUL-1999;	99US-0145085P.	PR	21-OCT-1999;	99US-0160814P.
PR	22-JUL-1999;	99US-0145087P.	PR	21-OCT-1999;	99US-0160815P.
PR	22-JUL-1999;	99US-0145089P.	PR	22-OCT-1999;	99US-0160980P.
PR	22-JUL-1999;	99US-0145192P.	PR	22-OCT-1999;	99US-0160981P.
PR	23-JUL-1999;	99US-0145145P.	PR	22-OCT-1999;	99US-0160989P.
PR	23-JUL-1999;	99US-0145218P.	PR	25-OCT-1999;	99US-0161404P.
PR	23-JUL-1999;	99US-0145224P.	PR	25-OCT-1999;	99US-0161405P.
PR	26-JUL-1999;	99US-0145276P.	PR	25-OCT-1999;	99US-0161406P.
PR	27-JUL-1999;	99US-0145913P.	PR	26-OCT-1999;	99US-0161359P.
PR	27-JUL-1999;	99US-0145918P.	PR	26-OCT-1999;	99US-0161360P.
PR	27-JUL-1999;	99US-0145919P.	PR	26-OCT-1999;	99US-0161361P.
PR	28-JUL-1999;	99US-0145951P.	PR	26-OCT-1999;	99US-0161920P.
PR	28-JUL-1999;	99US-0146386P.	PR	28-OCT-1999;	99US-0161932P.
PR	02-AUG-1999;	99US-0146388P.	PR	28-OCT-1999;	99US-0161933P.
PR	02-AUG-1999;	99US-0146389P.	PR	29-OCT-1999;	99US-0162142P.
PR	03-AUG-1999;	99US-0147038P.			
PR	04-AUG-1999;	99US-0147204P.			
PR	04-AUG-1999;	99US-0147302P.			
PR	05-AUG-1999;	99US-0147302P.			
PR	05-AUG-1999;	99US-0147192P.			
PR	05-AUG-1999;	99US-0147260P.			
PR	06-AUG-1999;	99US-0147303P.			
PR	06-AUG-1999;	99US-0147416P.			
PR	09-AUG-1999;	99US-0147493P.			
PR	09-AUG-1999;	99US-0147933P.			
PR	10-AUG-1999;	99US-0148171P.			
PR	11-AUG-1999;	99US-0148319P.			
PR	12-AUG-1999;	99US-0148341P.			
PR	13-AUG-1999;	99US-0148565P.			
PR	13-AUG-1999;	99US-0148684P.			
PR	16-AUG-1999;	99US-0149368P.			
PR	17-AUG-1999;	99US-0149175P.			
PR	18-AUG-1999;	99US-0149426P.			
PR	20-AUG-1999;	99US-0149723P.			
PR	20-AUG-1999;	99US-0149723P.			
PR	20-AUG-1999;	99US-0149929P.			
PR	23-AUG-1999;	99US-0149902P.			
PR	23-AUG-1999;	99US-0149930P.			
PR	25-AUG-1999;	99US-0150566P.			
PR	26-AUG-1999;	99US-0150884P.			
PR	27-AUG-1999;	99US-0151065P.			
PR	27-AUG-1999;	99US-0151066P.			
PR	27-AUG-1999;	99US-0151080P.			
PR	30-AUG-1999;	99US-0151303P.			
PR	31-AUG-1999;	99US-0151438P.			
PR	01-SEP-1999;	99US-0151930P.			
PR	07-SEP-1999;	99US-0152363P.			
PR	10-SEP-1999;	99US-0153070P.			
PR	13-SEP-1999;	99US-0153758P.			
PR	15-SEP-1999;	99US-0154018P.			
PR	16-SEP-1999;	99US-0154039P.			
PR	20-SEP-1999;	99US-0154779P.			
PR	22-SEP-1999;	99US-0155139P.			
PR	23-SEP-1999;	99US-0155486P.			
PR	24-SEP-1999;	99US-0155659P.			
PR	28-SEP-1999;	99US-0156458P.			
PR	29-SEP-1999;	99US-0156596P.			
PR	04-OCT-1999;	99US-0157117P.			
PR	05-OCT-1999;	99US-0157753P.			
PR	06-OCT-1999;	99US-0157865P.			
PR	07-OCT-1999;	99US-0158029P.			
PR	08-OCT-1999;	99US-0158232P.			
PR	12-OCT-1999;	99US-0158369P.			
PR	19-JUL-1999;	99US-0144331P.	PR	13-OCT-1999;	99US-0159293P.
PR	19-JUL-1999;	99US-0144332P.	PR	13-OCT-1999;	99US-0159294P.
PR	19-JUL-1999;	99US-0144333P.	PR	13-OCT-1999;	99US-0159295P.
PR	19-JUL-1999;	99US-0144334P.	PR	14-OCT-1999;	99US-0159329P.
PR	19-JUL-1999;	99US-0144335P.	PR	14-OCT-1999;	99US-0159330P.
PR	20-JUL-1999;	99US-0144336P.	PR	14-OCT-1999;	99US-0159331P.
PR	20-JUL-1999;	99US-0144337P.	PR	14-OCT-1999;	99US-0159637P.
PR	20-JUL-1999;	99US-0144338P.	PR	14-OCT-1999;	99US-0159638P.
PR	20-JUL-1999;	99US-0144884P.	PR	18-OCT-1999;	99US-0159584P.
PR	21-JUL-1999;	99US-0144814P.	PR	21-OCT-1999;	99US-0160741P.
PR	21-JUL-1999;	99US-0145086P.	PR	21-OCT-1999;	99US-0160767P.
PR	21-JUL-1999;	99US-0145088P.	PR	21-OCT-1999;	99US-0160770P.
PR	22-JUL-1999;	99US-0145085P.	PR	21-OCT-1999;	99US-0160814P.
PR	22-JUL-1999;	99US-0145087P.	PR	21-OCT-1999;	99US-0160815P.
PR	22-JUL-1999;	99US-0145089P.	PR	22-OCT-1999;	99US-0160980P.
PR	22-JUL-1999;	99US-0145192P.	PR	22-OCT-1999;	99US-0160981P.
PR	23-JUL-1999;	99US-0145145P.	PR	22-OCT-1999;	99US-0160989P.
PR	23-JUL-1999;	99US-0145218P.	PR	25-OCT-1999;	99US-0161404P.
PR	23-JUL-1999;	99US-0145224P.	PR	25-OCT-1999;	99US-0161405P.
PR	26-JUL-1999;	99US-0145276P.	PR	25-OCT-1999;	99US-0161406P.
PR	27-JUL-1999;	99US-0145913P.	PR	26-OCT-1999;	99US-0161359P.
PR	27-JUL-1999;	99US-0145918P.	PR	26-OCT-1999;	99US-0161360P.
PR	27-JUL-1999;	99US-0145919P.	PR	26-OCT-1999;	99US-0161361P.
PR	28-JUL-1999;	99US-0145951P.	PR	26-OCT-1999;	99US-0161920P.
PR	28-JUL-1999;	99US-0146386P.	PR	28-OCT-1999;	99US-0161932P.
PR	02-AUG-1999;	99US-0146388P.	PR	28-OCT-1999;	99US-0161933P.
PR	02-AUG-1999;	99US-0146389P.	PR	29-OCT-1999;	99US-0162142P.
PR	03-AUG-1999;	99US-0147038P.			
PR	04-AUG-1999;	99US-0147204P.			
PR	04-AUG-1999;	99US-0147302P.			
PR	05-AUG-1999;	99US-0147302P.			
PR	05-AUG-1999;	99US-0147192P.			
PR	05-AUG-1999;	99US-0147260P.			
PR	06-AUG-1999;	99US-0147303P.			
PR	06-AUG-1999;	99US-0147416P.			
PR	09-AUG-1999;	99US-0147493P.			
PR	09-AUG-1999;	99US-0147933P.			
PR	10-AUG-1999;	99US-0148171P.			
PR	11-AUG-1999;	99US-0148319P.			
PR	12-AUG-1999;	99US-0148341P.			
PR	13-AUG-1999;	99US-0148565P.			
PR	13-AUG-1999;	99US-0148684P.			
PR	16-AUG-1999;	99US-0149368P.			
PR	17-AUG-1999;	99US-0149175P.			
PR	18-AUG-1999;	99US-0149426P.			
PR	20-AUG-1999;	99US-0149723P.			
PR	20-AUG-1999;	99US-0149723P.			
PR	20-AUG-1999;	99US-0149929P.			
PR	23-AUG-1999;	99US-0149902P.			
PR	23-AUG-1999;	99US-0149930P.			
PR	25-AUG-1999;	99US-0150566P.			
PR	26-AUG-1999;	99US-0150884P.			
PR	27-AUG-1999;	99US-0151065P.			
PR	27-AUG-1999;	99US-0151066P.			
PR	27-AUG-1999;	99US-0151080P.			
PR	30-AUG-1999;	99US-0151303P.			
PR	31-AUG-1999;	99US-0151438P.			
PR	01-SEP-1999;	99US-0151930P.			
PR	07-SEP-1999;	99US-0152363P.			
PR	10-SEP-1999;	99US-0153070P.			
PR	13-SEP-1999;	99US-0153758P.			
PR	15-SEP-1999;	99US-0154018P.			
PR	16-SEP-1999;	99US-0154039P.			
PR	20-SEP-1999;	99US-0154779P.			
PR	22-SEP-1999;	99US-0155139P.			
PR	23-SEP-1999;	99US-0155486P.			
PR	24-SEP-1999;	99US-0155659P.			
PR	28-SEP-1999;	99US-0156458P.			
PR	29-SEP-1999;	99US-0156596P.			
PR	04-OCT-1999;	99US-0157117P.			
PR	05-OCT-1999;	99US-0157753P.			
PR	06-OCT-1999;	99US-0157865P.			
PR	07-OCT-1999;	99US-0158029P.			
PR	08-OCT-1999;	99US-0158232P.			
PR	12-OCT-1999;	99US-0158369P.			

Query Match 28.8%; Score 886; DB 3; Length 475;

Best Local Similarity 39.9%; Pred. No. 1.3e-80;

Matches 195; Conservative 89; Mismatches 161; Indels 44; Gaps 10;

QY	69	MHDWLVGL-----SRHRTVTVDMPFTS-----YTIADPNNVHVIVKTNFTNPKGVYR	119
DB	1	MHDEIADNLRMCGGYTCIPPIFLAKQGHVTVTCDPKNLEHLKTRFDNPKGSPSWQ	60
QY	120	SYMDVLLDGGIFNADGELWKRQKRTAFSEFASKNLRDFA-IVFREYSLKLSGLSQAASK	178
DB	61	SVFHDLLDGGIFNDSGDDTNRFORKTALEFFTRTLRQAMAEWVDRAIKNLVPILESARS	120
QY	179	AGKVVDMOELYMRTLDSICKVGFGEIGTSLPDPENSFAQAFDAANIITLRFIDP--	236
DB	121	RAEPIDLDQVLLRLTDFDNICGLTFGKOPRTLSPEFPNGFAVAFDGAETALQRFIMPEF	180
QY	237	LWRIKRPFHVGSEALLAQSIKLVDBFTYSVIRKAEIIEVVRASGKQEKMKHDIILSRTE	296
DB	181	IWKIRKWLRLGLEDDMSRSISHDVNYLSEIINTRKLELLGQQDQGRH---DDLRSRMK	237
QY	297	LGEAGDDGGFGDDKSLRDVVLNFVIAGRDTTATLSWFTHMAMSHPDVAEKLRRELCAF	356
DB	238	KKES-----YSDKYLKVALNFILAGRTSSVAMSWFFWLVSINPRVEKIINEIC--	288
QY	357	EAERAREGVTLVLCGGADADDKAPAAVAFAG-LITYDSLGLKLVYLHACVTEILRLYP	415
DB	289	-----TILI-----KTRDTNVSKWTDPLTFDEIDQLVYLKAALSETLRLYP	330
QY	416	AVPQDPKGILEDVLPDGTQVVRAGMTVYVSYSGRMVYNGPDAAASPRPERWINEGCAF	475
DB	331	SYPEDSKFVANDVLPDGTFFVPSGNSVTYSYSGRMKFTWGEDCLEFEPKRWL-EESRD	389
QY	476	RNASPFKFTAFQAGPRICLGKDSAYLOMKMALAIFREYSPFLLEGHPVQYRMMTILSMA	535
DB	390	EKCNOYKFAVFNAGPRICLGKDLAYLOMKSIATASILLRHRLTVAPGHRVQKMSUTLPMK	449
QY	536	HGLKVRVSR	544
DB	450	FGLKMDVHK	458

RESULT 12  
AAG54205  
ID AAG54205 standard; protein; 465 AA.

XX AC AAG54205;  
XX DT 18-OCT-2000 (first entry)  
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 69089.  
XX KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
XX OS Arabidopsis thaliana.  
XX PN EP1033405-A2.  
XX PD 06-SEP-2000.  
XX PF 25-FEB-2000; 2000EP-00301439.  
XX PR 25-FEB-1999; 99US-0121825P.  
XX PR 05-MAR-1999; 99US-0123180P.  
XX PR 09-MAR-1999; 99US-0123548P.  
XX PR 23-MAR-1999; 99US-0125788P.  
XX PR 25-MAR-1999; 99US-0126264P.  
XX PR 29-MAR-1999; 99US-0126785P.  
XX PR 01-APR-1999; 99US-0127462P.  
XX PR 06-APR-1999; 99US-0128234P.  
XX PR 08-APR-1999; 99US-0128714P.  
XX PR 16-APR-1999; 99US-0129845P.  
XX PR 19-APR-1999; 99US-0130077P.  
XX PR 21-APR-1999; 99US-0130449P.  
XX PR 23-APR-1999; 99US-0130510P.  
XX PR 23-APR-1999; 99US-0130891P.  
XX PR 28-APR-1999; 99US-0131449P.  
XX PR 30-APR-1999; 99US-0132048P.  
XX PR 30-APR-1999; 99US-0132407P.  
XX PR 04-MAY-1999; 99US-0132484P.  
XX PR 05-MAY-1999; 99US-0132485P.  
XX PR 06-MAY-1999; 99US-0132486P.  
XX PR 06-MAY-1999; 99US-0132487P.  
XX PR 07-MAY-1999; 99US-0132863P.  
XX PR 11-MAY-1999; 99US-0134256P.  
XX PR 14-MAY-1999; 99US-0134218P.  
XX PR 14-MAY-1999; 99US-0134219P.  
XX PR 14-MAY-1999; 99US-0134221P.  
XX PR 14-MAY-1999; 99US-0134370P.  
XX PR 18-MAY-1999; 99US-0134768P.  
XX PR 19-MAY-1999; 99US-0134941P.  
XX PR 20-MAY-1999; 99US-0135124P.  
XX PR 21-MAY-1999; 99US-0135353P.  
XX PR 24-MAY-1999; 99US-0135629P.  
XX PR 25-MAY-1999; 99US-0136021P.  
XX PR 27-MAY-1999; 99US-0136392P.  
XX PR 28-MAY-1999; 99US-0136782P.  
XX PR 01-JUN-1999; 99US-0137222P.  
XX PR 03-JUN-1999; 99US-0137528P.  
XX PR 04-JUN-1999; 99US-0137502P.  
XX PR 07-JUN-1999; 99US-0137724P.  
XX PR 08-JUN-1999; 99US-0138094P.  
XX PR 10-JUN-1999; 99US-0138540P.  
XX PR 10-JUN-1999; 99US-0138847P.  
XX PR 14-JUN-1999; 99US-0139119P.  
XX PR 16-JUN-1999; 99US-0139452P.  
XX PR 16-JUN-1999; 99US-0139453P.  
XX PR 17-JUN-1999; 99US-0139492P.  
XX PR 18-JUN-1999; 99US-0139454P.  
XX PR 18-JUN-1999; 99US-0139455P.  
XX PR 18-JUN-1999; 99US-0139456P.  
XX PR 18-JUN-1999; 99US-0139457P.  
XX PR 18-JUN-1999; 99US-0139458P.  
XX PR 18-JUN-1999; 99US-0139459P.  
XX PR 18-JUN-1999; 99US-0139460P.  
XX PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 18-JUL-1999; 99US-0144325P.  
PR 18-JUL-1999; 99US-0144331P.  
PR 18-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 09-AUG-1999; 99US-0148171P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.



us-10-021-657-2.rag

Mon Mar 8 10:13:27 2004

PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 21-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147202P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161992P.  
PR 28-OCT-1999; 99US-0161993P.

```
PR 29-OCT-1999; 99US-0162142P.
Query Match 27.5%; Score 847; DB 3; Length 508;
Best Local Similarity 36.2%; Pred. No. 1.4e-76;
Matches 195; Conservative 90; Mismatches 188; Indels 66; Gaps 12;
QY 27 IALLVVLVSWIIVQWLSLKKQ-----GPRSWPVIGATVQLRNHYHRHMDLVGY--L 77
DB 6 ILILFIIIGFVLFISPTTLAKTASPNSPIISYPLGISLFSNKNRHLQWYTDLRL 65
QY 78 SRHRTVVDMPSTSYNI-ADPNVVEHLKTNFTNPKGIVYRSYMDVLLGGIENADGE 136
DB 66 SPSQITVLLFORRTIITANPENVEHLKTNFTNPKGKPTDLLGDLGGIENSDGE 125
QY 137 LMKQKQKTASFEPFASKNLDFSAIVR-FYSKLKSLGILSOAKGKVKVDMQELYMRTLD 195
DB 126 LWSQKQLASHFTMSLREFTFEILREEVQNRLIPVLSAVDCGETVDFQEVLFKFAFD 185
QY 196 STCKVGFVGEIGTSLPDLFENSPAQAFDAANIITLRFIDP---LWRIKRFHVHVGSEALL 252
DB 186 VVCKVSLGWDPCDCLDTRVPVELVKAFDVAAEISARRATEPVYAVWVKRFLNVGSEKRL 245
QY 253 AOSIKLVDEFTYSVIRRRKAEIVEVRASGKQKMKHDIILSRFTELGEAGDGGGFGDDKS 312
DB 246 REAKTVHLSVSEIIIAKKSL---DIGDVSDKQDLSRFLAAGHG-----EEA 292
QY 313 LRDVILNFVIAGDRTTATTLISWFTNMMSHPOVAELRLRELCAFEARAREEGVTLVLCG 372
DB 293 VRDSVISFIMAGDRTTSAAMTFLWLLSQNDVETKILDEL-----RNKG----- 337
QY 373 GADADOKAPARVAQFAGLLTYDSLGLKLVYLHACVTELRLYPAVQDPKGILEDVLPD 432
DB 338 -----SLGLGFDLREMSYTKALCEAMRLYPPVAVDSKHAANDLILPD 361
QY 433 GTRVAGGVTVYVPSMGRMEYNWGPDAASFRPERMINEDGAF-----RNASPPKFTAF 486
DB 382 GTPLKKGDKVTYFPYNGRMEKVKWGDWDEFKPNRWFEEPSYGTKEPVLSVSSFKFVF 441
QY 487 QAGPRICLGKDSAYLQMKMAL-AILFRYFRLLEGHVQYRMVMTLSMAHGLKVRVSR 544
DB 442 QAGPRVICGEMAFQTKMYVGVSVLSRFKIPVCNNRPVFPVPLIT-AHMAAGGLKVKIKR 499

RESULT 14
AAG18700
ID AAG18700 standard; protein; 510 AA.
XX
AC AAG18700;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 20214.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
DD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
XX 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0128845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142377P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
```

PR	13-OCT-1999;	99US-0159295P.
PR	14-OCT-1999;	99US-0159329P.
PR	14-OCT-1999;	99US-0159330P.
PR	14-OCT-1999;	99US-0159331P.
PR	14-OCT-1999;	99US-0159637P.
PR	14-OCT-1999;	99US-0159638P.
PR	18-OCT-1999;	99US-0159584P.
PR	21-OCT-1999;	99US-0160741P.
PR	21-OCT-1999;	99US-0160767P.
PR	21-OCT-1999;	99US-0160768P.
PR	21-OCT-1999;	99US-0160770P.
PR	21-OCT-1999;	99US-0160814P.
PR	21-OCT-1999;	99US-0160815P.
PR	22-OCT-1999;	99US-0160380P.
PR	22-OCT-1999;	99US-0160381P.
PR	22-OCT-1999;	99US-0160389P.
PR	25-OCT-1999;	99US-0161404P.
PR	25-OCT-1999;	99US-0161405P.
PR	25-OCT-1999;	99US-0161406P.
PR	25-OCT-1999;	99US-0161359P.
PR	26-OCT-1999;	99US-0161360P.
PR	26-OCT-1999;	99US-0161361P.
PR	28-OCT-1999;	99US-0161920P.
PR	28-OCT-1999;	99US-0161922P.
PR	28-OCT-1999;	99US-0161933P.
PR	29-OCT-1999;	99US-0162142P.
Query Match 27.5%; Score 847; DB 3; Length 510;		
Best Local Similarity 36.2%; Pred. No. 1.4e-76;		
Matches 195; Conservative 90; Mismatches 188; Indels 66; Gaps 12;		
Qy	27	IALLVVLVSLVQMSLRKQK-----GPRSWPVIGATVEQLNRYHRMHDVLGY--L 77
Db	8	ILIFPIIGFVLIFSPPTKTKAKTASPSNPISTPLIGSILSNKNRHLQWYTDLLRL 67
Qy	78	SRHRTVDMPTFTSYTI-ADPVNVEHLKTNFTNPKGIVYASYMDVLGDLGNADGE 136
Db	68	SPSQTIIVDLLFGERTITANPENVEHLKTNFTNPKGPKFTDLLGDLGGGIFNSDGE 127
Qy	137	LWRKQRTASFEFASKNLRDPSAIVFR-EYSLKLSGLSQASKAGKVVDQELMYRTLD 195
Db	128	LWSSQRKLASHEFTMRSLEFTEILREEVQNRLLIPVLSSAVDCGETVDFQEVLRKFAFD 187
Qy	196	SICKVGFGEVIGTILSPDLPENSFAQAFDAANIITLRFIDP---LWRIKFFHVGSALL 252
Db	188	VVCKVSLGWDPDCLDLTRPVPELVKADFVAEASARRATEFVAVWKVKFLNVSSEKRL 247
Qy	253	AQSIKLVDEFTYSVIRRRKAEIVEVRASGKQEKMKHDILSRFIELGSEAGDDGGFGDDKS 312
Db	248	REAIKTVHLSVSEIIRAKKSL---DIGGDVSDKQDLLSRFLAAGHG-----EEA 294
Qy	313	LRDVLNFTVAGRTTATTLNFWTHMAMSHPDVAEKLRRELCAFEAREEGVTILVCG 372
Db	295	VRDSVISFIMAGROTTSAAMTWLFWLLSQNDVETKILDEL-----RNKG----- 339
Qy	373	GADADDKAFARVAQAFAGLTLYDSLGLVYLHACVTETTLKLYPAVPODPKGILEDVLPD 432
Db	340	-----SLIGGFEDLRMSYTKACLCEAMRLYPFVANDSKHAANDDILPD 383
Qy	433	GTKVRAGGMVTYFYPYSGRMENYNGWPDAAASFRPRERWINEGAF-----RNASPKFTAF 486
Db	384	GTPLKKGDKVTYFYPYSGRMENYNGWPDAAASFRPRERWINEGAF-----RNASPKFTAF 443
Qy	487	QAGPRICLGKDSAYLQNMAL-AILFRFYFRLEGHVPQYRMWMTILSMAHGLKVRYSR 544
Db	444	QAGPRVCIKEMAFQMKYVGVSVLSRFKIIPVCNRPFPVPLLT-AHMAGGLKVKIKR 501
RESULT 15		
AAG18699		
ID AAG18699 standard; protein; 528 AA.		
XX		
AC AAG18699;		

Mon Mar 8 10:13:27 2004

us-10-021-657-2.rag

XX	17-OCT-2000 (first entry)	PR	18-JUN-1999;	99US-0139750P.
DT		PR	18-JUN-1999;	99US-0139763P.
XX	Arabidopsis thaliana protein fragment SEQ ID NO: 20213.	PR	21-JUN-1999;	99US-0139817P.
DE		PR	22-JUN-1999;	99US-0139899P.
XX		PR	23-JUN-1999;	99US-0140353P.
XX	Protein identification; signal transduction pathway; metabolic pathway;	PR	23-JUN-1999;	99US-0140354P.
KW	hybridisation assay; genetic mapping; gene expression control; promoter;	PR	24-JUN-1999;	99US-0140695P.
KW	termination sequence.	PR	28-JUN-1999;	99US-0140823P.
XX		PR	29-JUN-1999;	99US-0140991P.
OS	Arabidopsis thaliana.	PR	30-JUN-1999;	99US-0141287P.
XX		PR	01-JUL-1999;	99US-0141842P.
XX	EP1033405-A2.	PR	01-JUL-1999;	99US-0142154P.
XX		PR	02-JUL-1999;	99US-0142055P.
XX	06-SEP-2000.	PR	06-JUL-1999;	99US-0142390P.
XX		PR	08-JUL-1999;	99US-0142803P.
XX		PR	09-JUL-1999;	99US-0142920P.
XX		PR	12-JUL-1999;	99US-0142977P.
XX	25-FEB-2000; 2000EP-00301439.	PR	13-JUL-1999;	99US-0143542P.
XX		PR	14-JUL-1999;	99US-0143624P.
PR	25-FEB-1999;	PR	15-JUL-1999;	99US-0144005P.
PR	05-MAR-1999;	PR	16-JUL-1999;	99US-0144085P.
PR	09-MAR-1999;	PR	16-JUL-1999;	99US-0144086P.
PR	23-MAR-1999;	PR	19-JUL-1999;	99US-0144325P.
PR	25-MAR-1999;	PR	19-JUL-1999;	99US-0144331P.
PR	29-MAR-1999;	PR	19-JUL-1999;	99US-0144332P.
PR	01-APR-1999;	PR	19-JUL-1999;	99US-0144333P.
PR	06-APR-1999;	PR	19-JUL-1999;	99US-0144334P.
PR	08-APR-1999;	PR	19-JUL-1999;	99US-0144335P.
PR	16-APR-1999;	PR	20-JUL-1999;	99US-0144332P.
PR	19-APR-1999;	PR	20-JUL-1999;	99US-0144804P.
PR	21-APR-1999;	PR	21-JUL-1999;	99US-0144814P.
PR	23-APR-1999;	PR	21-JUL-1999;	99US-0145086P.
PR	28-APR-1999;	PR	21-JUL-1999;	99US-0145088P.
PR	30-APR-1999;	PR	22-JUL-1999;	99US-0145085P.
PR	04-MAY-1999;	PR	22-JUL-1999;	99US-0145087P.
PR	05-MAY-1999;	PR	22-JUL-1999;	99US-0145192P.
PR	06-MAY-1999;	PR	22-JUL-1999;	99US-0145193P.
PR	07-MAY-1999;	PR	23-JUL-1999;	99US-0145145P.
PR	11-MAY-1999;	PR	23-JUL-1999;	99US-0145218P.
PR	14-MAY-1999;	PR	26-JUL-1999;	99US-0145224P.
PR	14-MAY-1999;	PR	26-JUL-1999;	99US-0145276P.
PR	14-MAY-1999;	PR	27-JUL-1999;	99US-0145913P.
PR	14-MAY-1999;	PR	27-JUL-1999;	99US-0145918P.
PR	14-MAY-1999;	PR	27-JUL-1999;	99US-0145919P.
PR	18-MAY-1999;	PR	28-JUL-1999;	99US-0145951P.
PR	19-MAY-1999;	PR	28-JUL-1999;	99US-0146386P.
PR	20-MAY-1999;	PR	02-AUG-1999;	99US-0146388P.
PR	21-MAY-1999;	PR	02-AUG-1999;	99US-0146389P.
PR	24-MAY-1999;	PR	03-AUG-1999;	99US-0147038P.
PR	25-MAY-1999;	PR	04-AUG-1999;	99US-0147204P.
PR	27-MAY-1999;	PR	04-AUG-1999;	99US-0147302P.
PR	28-MAY-1999;	PR	05-AUG-1999;	99US-0147192P.
PR	01-JUN-1999;	PR	05-AUG-1999;	99US-0147260P.
PR	03-JUN-1999;	PR	06-AUG-1999;	99US-0147303P.
PR	04-JUN-1999;	PR	06-AUG-1999;	99US-0147416P.
PR	07-JUN-1999;	PR	09-AUG-1999;	99US-0147493P.
PR	08-JUN-1999;	PR	09-AUG-1999;	99US-0147935P.
PR	10-JUN-1999;	PR	10-AUG-1999;	99US-0148171P.
PR	10-JUN-1999;	PR	11-AUG-1999;	99US-0148319P.
PR	14-JUN-1999;	PR	12-AUG-1999;	99US-0148341P.
PR	16-JUN-1999;	PR	13-AUG-1999;	99US-0148565P.
PR	16-JUN-1999;	PR	13-AUG-1999;	99US-0148684P.
PR	17-JUN-1999;	PR	16-AUG-1999;	99US-0149368P.
PR	18-JUN-1999;	PR	17-AUG-1999;	99US-0149175P.
PR	18-JUN-1999;	PR	18-AUG-1999;	99US-0149426P.
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149722P.
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149723P.
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149929P.
PR	18-JUN-1999;	PR	23-AUG-1999;	99US-0149902P.
PR	18-JUN-1999;	PR	23-AUG-1999;	99US-0149930P.
PR	18-JUN-1999;	PR	25-AUG-1999;	99US-0150566P.
PR	18-JUN-1999;	PR	26-AUG-1999;	99US-0150884P.
PR	18-JUN-1999;	PR		

PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0156559P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161922P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.  
PR 29-OCT-1999; 99US-0162142P.  
Query Match 27.5%; Score 847; DB 3; Length 528;  
Best Local Similarity 36.2%; Pred No. 1.5e-76;  
Matches 195; Conservative 90; Mismatches 188; Indels 66; Gaps 12;  
QY 27 IALLVVLVSWILVQWLSLRKQ-----GPRSWPVIGATVEQLRNYRHMEDWLGY--L 77  
DB 26 ILILFPIIGVLIFSPPTKLTAKTASPSNFI SYPLIGSILSFNKNRHLQLQWYTDLLRL 85  
QY 78 SRHRTVTVMPPFTSYI-ADPNVVEHLKTNFTNPKGIVRSYMDVLGDGIFNADGE 136  
DB 86 SPSQTIIVDLLFGRRTIITANPENVEHLKTNFTNPKGKPFETDLLGDLGGGIFNSDGE 145  
QY 137 LWRKQRTASPEKSLRDFSAIVR-EYSLKLSGILSOAKAGKVVDVMDQELVMRTLD 195  
DB 146 LWSQRKLASHEFTMRSLREFTFELREEVQNRLLI PVLSAVDCGETVDQEVLRKFAFD 205  
QY 196 SICKVGFGEIGTILSPDLPENSAQAFDAANIITLRFIDP---LWRIKFFHVHVSALL 252  
DB 206 VVCKVSLGWDPCDLDTFRFVPELVKAFDVAEISARRATEPVYAVKVKRFLNVGSEKRL 265  
QY 253 AQSLKLVDEFTYSVIRRKAEIIEVRASGKOEKMKHILSRFIELGEAGDDGGGFGDDKS 312

DB 266 REAKTVHLSVSEIRAKKSL-----DIGGDVSDKQDLSRFLAAGHG-----EEA 312  
QY 313 LRDVVLNFIAGRDTTATLTSWFTHMAMSHPDVAEKLELRELCAPAEARBEVTLVLCG 372  
DB 313 VRDSVISFIMAGRDTTSAAMTWLFWLLSQNDVETKILDEL-----RNKG----- 357  
QY 373 GADADDKAFARVAQFAGLLTYDSILGLVYLHACVETETRLYPVPQDPKGLIEDDVLDP 432  
DB 358 -----SLGLGFEDLREMSYTKACICEAMRLYPVAVWDSKHAANDDILPD 401  
QY 433 GTKVRAGGMVTVPYSMGRMEYNWGPDAASFRPERWINEGDAF-----RNASPEKFTAF 486  
DB 402 GTPLKKGDKVTFYFPYGMGMEKWKWDDEKPNRWFEEPSYCTKPVLSVSSFKFPVF 461  
QY 487 QAGPRICLGKDSAYLQMKAL-AILFRFYSRLLEGHVQCYRMMTILSMAGHLKVVR 544  
DB 462 QAGPRVCIGKEMAFQMKYVGVSVLSRFLKIIIPVCNNRPVFPVPLLT-AHMAGGLKYIKR 519

Search completed: March 2, 2004, 15:27:43  
Job time : 63 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2004, 15:25:49 ; Search time 21 Seconds  
(without alignments)  
2693.363 Million cell updates/sec

Title: US-10-021-657-2  
Perfect score: 3079  
Sequence: 1 EFGTREHLTPATPSFFPL.....ICMHACKGRWVSLVAVLKP 588

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues  
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	994.5	32.3	553	T01535	probable cytochrom
2	977.5	31.7	511	T00864	cytochrome P450 ho
3	971.5	31.6	537	T02450	probable cytochrom
4	970	31.5	554	G86146	hypothetical prote
5	969.5	31.5	523	B96662	probable cytochrom
6	907.5	29.5	490	T00404	cytochrome P450-11
7	904.5	29.4	550	T50510	cytochrome P450-11
8	879.5	28.6	499	T49064	cytochrome P450 cy
9	876.5	28.5	513	JC5965	protein F2119.20
10	837	27.2	522	B86339	cytochrome P450-11
11	823	26.7	506	T46196	cytochrome P450-11
12	788	25.6	529	G86265	F3F19.17 protein -
13	795.5	25.5	514	T08014	cytochrome P450 cy
14	780	25.3	519	B86265	hypothetical prote
15	773	25.1	495	G84675	probable cytochrom
16	744.5	24.2	516	T09364	cytochrome P450 ho
17	744	24.2	516	T00514	cytochrome P450 ho
18	738.5	24.0	510	F84606	probable cytochrom
19	730	23.7	479	T09365	cytochrome P450 ho
20	727.5	23.6	506	T04628	cytochrome P450 ho
21	718.5	23.3	497	G96611	probable cytochrom
22	699	22.7	508	T09367	cytochrome P450 ho
23	682	22.2	503	T02357	cytochrome P450 ho
24	675	21.9	512	C96517	hypothetical prote
25	671	21.8	520	B96517	hypothetical prote
26	650	21.1	469	T09366	cytochrome P450 ho
27	619.5	20.1	480	T48311	cytochrome P450 52
28	427	13.9	519	S69988	unspecific monooxy
29	411	13.3	519	S69989	unspecific monooxy

30	408	13.3	593	2	F86441	probable cytochrom
31	407.5	13.2	523	1	O4CKA3	cytochrome P450 52
32	402	13.1	524	1	S22972	cytochrome P450 52
33	400.5	13.0	523	1	J01039	cytochrome P450 52
34	399	13.0	526	1	A40576	cytochrome P450 AL
35	392	12.7	538	1	B40576	cytochrome P450 AL
36	387	12.6	519	1	J50726	cytochrome P450 AL
37	386.5	12.6	524	2	JC7594	cytochrome P450 en
38	386.5	12.6	524	2	JC7598	cytochrome P450 en
39	385	12.5	538	1	O4CKA4	cytochrome P450 52
40	384.5	12.5	519	1	J50725	cytochrome P450 AL
41	383.5	12.5	520	2	T24780	hypothetical prote
42	381	12.4	522	1	JT0980	cytochrome P450 52
43	380.5	12.4	518	2	T24783	hypothetical prote
44	374	12.1	544	2	T27750	hypothetical prote
45	371	12.0	520	1	A46661	leukotriene B4 one

ALIGNMENTS

RESULT 1

T01535  
probable cytochrome P450 A IG005110.21 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cross)  
C>Date: 19-Feb-1999 #sequence\_revision 19-Feb-1999 #text\_change 28-Jul-2000  
C/Accession: T01535  
R:Andrews, S.  
submitted to the EMBL Data Library, July 1997  
A:Description: The sequence of A. thaliana IG005110.  
A:Reference number: Z14347  
A/Accession: T01535  
A/Status: translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-553 <AND>  
A/Cross-references: EMBL:AF013293; NID:g2252823; PID:g2252844  
A/Experimental source: Cultivar Columbia  
C/Genetics:  
A/Map position: 4  
A/Introns: 141/2  
A/Note: A IG005110.21  
A/Superfamily: Candida cytochrome P450 52A3; cytochrome P450 homology  
F:304-481/Domain: cytochrome P450 homology <P45>

Query Match	32.3%	Score	994.5	DB	2	Length	553
Best Local Similarity	41.0%	Pred. No.	3.9e-89	Indels	51	Gaps	15
Matches	226	Conservative	92	Mismatches	182		
QY	29	LLLVLS--WILVQWLSLRKQKGRSPWPVIGATVEQLRNYHRMHDWLVGYL----	SRHRT	82			
Db	8	LLVAVAAVWLFQRLS--RWLKGPRVWVVLGSLGLEQDRMHDWITENLRACGGTYQT	66				
QY	83	VTVDMPFTS-----YTIADPVNVVHVLKTNFTNYPKGIVYRMYMDVLLGGGIFNADGEL	137				
Db	67	CICAVFPLAKQGLVTVTCDEKNIEMHLKTRFDNPKGPTQAVFVDFLQGGIFNSDGD	126				
QY	138	WRQKRTASPEFASKNLRD--FSAIVFREYSLSKLSILSQASKAGKVVDMLGYRWTLDS	196				
Db	127	WLFQRTAALEFTRTLRQANGWVARGIKLFCPILETQNNVPEVDQLLIRLITFDN	186				
QY	197	ICKVGFGEVIGTSPDLNPENSAQAFDAANIITLAFIDP--LWRIKRFHVSGSEALLAQ	254				
Db	187	ICGLAFQKTRTCAPGLPENGASAFDRATEASLQRFILFEFLWRLKKWLGLEVLSLR	246				
QY	255	SIKLVDFTYSVIRRRKAEIVFRASQKQMKH--DILSRFIELGEAGDGGGGGDDKSL	313				
Db	247	SLGIDGYLDVAVINTRKQLLSQRESGVQ---RHDDLRSRFMKKKDQS-----YSETFL	297				
QY	314	RDVVLNVFIAGRDTATLTSWFTHMASHDDVAEKLRRELCAFEAREEGVTLVLCGG	373				
Db	298	RHVALNFILAGRDTSSVALSFFWFLITHTPTVEDKIVREICS-----VLIETRG	346				
QY	374	ADADDKAFARVAQFAGLITVDSLGKLVHLHACVTETRLYPVAVPQDPKGILEDDVLPDG	433				

```

Db      347 TDVS-----SWTAEPLEFDEVDRLVYLKAALSETRLVPSVPDSKHWNDLILPDG 398
Qy      434 TKVAGMVIYVYSGRMENWCPDAASPRPERWIN-EDGAFNAPSPFKETAFQAGPRI 492
Db      399 TFPAGSSVTYSIYAAGRMKSTWGEDCLEFKPERWISPDGKFNHQQYRFVAFNAGPRI 458
Qy      493 CLGKDSAYLQWK-MALAILFRFYFRLLGHPVQYRMWMTILSMAHGLKVRVSRAVCHGDL 551
Db      459 CLGKDLAYLQWKTTAAVLLR-HLTVAPGHKVEQKMSLTLFMKNGLLVNVHKE---DL 513
Qy      552 DMDIVPLNPRQ 562
Db      514 EVMKSLVPKE 524

RESULT 2
T00864
cytochrome P450 homolog F17K2.4 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 16-Feb-2001
C:Accession: T00864; D84891
R:Rounsley, S.D.; Kaul, X.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes,
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Niernan, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: D84891
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-511 <ROU>
A:Cross-references: EMBL:AC003680; NID:g2979540; PID:g2979544
A:Experimental source: cultivar Columbia
R:Lin, X.; Kaul, X.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Niernan, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: D84891
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-511 <STO>
A:Cross-references: GB:AE002093; NID:g2979544; PIDN:AAC06153.1; GSPDB:GNO00139
C:Genetics:
A:Gene: At2g45510; F17K2.4
A:Map position: 2
A:Introns: 174/3; 269/3; 382/3; 449/3
C:Superfamily: Candida cytochrome P450 52A3; cytochrome P450 homology
F:302-477/Domain: cytochrome P450 homology <P45>

Query Match      31.7%; Score 977.5; DB 2; Length 511;
Best Local Similarity 41.7%; Pred. No. 7.2e-68;
Matches 215; Conservative 85; Mismatches 183; Indels 33; Gaps 7;

Qy      27 IALLVLSWILVQWSLRKQKPRSPVIGATVEQLRNHYRHMHWLGYLSRHRIVTVYD 86
Db      15 IFVLCTIYLMIRIFPTGRNDKRYAPVHATVFDLLFHSDELYDYEIAREKPTIRFL 74
Qy      87 MPFTSYTIADPNVVEHLKTNFTNPKGIYVRSYMDVLLGDIENADGELWRKQRTAS 146
Db      75 SPQGSILTPADPNVVEHLKTNFTNPKGIYVRSYMDVLLGDIENADGELWRKQRTAS 134
Qy      147 FEFASKNLRDFSIAVREYSLKSLGSLQSAGKAGVVDVMDQELYMRTLSICKVGFGEI 206
Db      135 FEFSTRVLRDFSCSVFRNASKLVGVSEFALSKAFADQDLMLRCTLDSIPKVGFGVEL 194
Qy      207 GTLSPLPE-NSFAQAFDAANIITIRFDPLWRKFRFFHVGSEALLAQSIKLVDFTYS 265
Db      195 KCLDGFSGEQEFMEAFDEGNVATSSRFIDPLWKLWFFNIGSQSKLKSIATIDKFVS 254
Qy      266 VIRRRKAEIVVEASGQEKWKHDIILSRFIELGEAGDGGGFGDKSLRDVILNFVIAGR 325
Db      255 LITTKREL-----AKEQNTVVRDILSRFIESEKDPEN---MNDKYLRLDIILNFVIAGK 307

```

```

Qy      326 DTTATTLSTWFTWAMSHPDVAEKLREL--CAFEARAREEGVTLLVCGGADADDKAFAA 383
Db      308 DTTAALLSNFLVLMCKNPLVQEKIVQIRDVTFESHEKTTD----- 347
Qy      384 RVAQFAGLLTYDSLGLVYLHACVTETRLALYPAVPODPKGILEDDEDDVLPDQTKVRAGGMVT 443
Db      348 -VNGFVESINEEDALDEHMLHAALSETRLRYPVPVDMRCAENDVLPDGHVRVSKGDIY 406
Qy      444 YVYSYSGRMENWCPDAASPRPERWINEDGAFNAPSPFKETAFQAGPRICLGDKDSAYLQW 503
Db      407 YIAYAMGRMTYIINGQDAEBFKPERML-KOGLFQPSFPFKFISFHAGRGICLGDKFAYRQM 465
Qy      504 KMALAILFRFYSFRLL-EGHPVQYRMWMTILSMAHGL 538
Db      466 KIVSMALLHFRFKMADENSKVYKRMMLTLHVDGGL 501

RESULT 3
T02450
probable cytochrome P450 F4I18.5 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 16-Feb-2001
C:Accession: T02450; B84897
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul,
submitted to the EMBL Data Library, August 1998
A:Description: Arabidopsis thaliana chromosome II BAC F4I18 genomic sequence.
A:Reference number: Z14674
A:Accession: T02450
A>Status: translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-537 <ROU>
A:Cross-references: EMBL:AC004665; NID:g3386593; PID:g3386598
A:Experimental source: cultivar Columbia
R:Lin, X.; Kaul, X.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Niernan, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: B84897
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-537 <STO>
A:Cross-references: GB:AE002093; NID:g6598466; PIDN:AAC62873.2; GSPDB:GNO0139
C:Genetics:
A:Gene: At2g45970; F4I18.5
A:Map position: 2
C:Superfamily: Candida cytochrome P450 52A3; cytochrome P450 homology
F:302-480/Domain: cytochrome P450 homology <P45>

Query Match      31.6%; Score 971.5; DB 2; Length 537;
Best Local Similarity 40.9%; Pred. No. 2.3e-67;
Matches 217; Conservative 87; Mismatches 182; Indels 45; Gaps 12;

Qy      27 IALLVLSWILVQWSLRKQKPRSPVIGATVEQLRNHYRHMHWLGYL-----SRHRT 82
Db      7 LMLSAITAYFLWLTFISRCCLKPRVWILGSLPGLIENCERHMDWISDNLACSGTQYT 66
Qy      83 VTVDMPFTS-----YTIADPNVVEHLKTNFTNPKGIYVRSYMDVLLGDIENADGEL 137
Db      67 CICALPFLAKKQGLTVITCDPRNLEHLKRNFDNYPKGTQWQAVPHDLLGQIGFNSDGT 126
Qy      138 WKQKRTASFEFPASKNLRDFS-IVFREYSLKSLGSLQSAGKAGVVDVMDQELYMRTLDS 196
Db      127 WLFQRTAALEPTTTLRQAMARWVNRALKRLFLILENARLGSEPIQLDILLLETFDN 186
Qy      197 ICKVGFGEIIGTSLDPLNSFAQAFDAANIITIRFDPLWRKFRFFHVGSEALLAQ 254
Db      187 ICGLFGDKPRTCAPGLPVNTFAVADPRATEASLQRFILPEILWFKRWLRGLGLEVSLTR 246
Qy      255 SIKLVDFTYSVIRRRKAEIVVEASGQEKWKHDIILSRFIELGEAGDGGGFGDKSLR 314

```

Db 247 SLVQVNDYSEIITTKKEEMTOHNGKH---DULLSRPIKKES-----YSDETIQ 296

QY 315 DVVNLNFIAGROTTATTLTWSFTTHAMSHPDVAEKLRLRELCAPAEARAREEGVTLVLCGA 374

Db 297 RVALNFIILAGDRTSSVALSWFFLIITQHPAIEDKILREICTVLVE-TRGDDVALM----- 350

QY 375 DADKFAARVAQFAGLLTYDSLGLKLYLHACVTEITLRLYPVAPQDPKGLLEDVLPDGT 434

Db 351 -TDEP-----LSCEELDRVLKALSETLRLYPSVPDSKRAVKDDVLPDGT 397

QY 435 KVRAGGVTVPYSGRMENWGPDAASFRPERWINE--GAFRNASPFKFTAFQAGPRI 492

Db 398 FVPAGSSITTYISAGRMKSTWGEDCLEFPERWISQSDGGRFINHDPFKFVAFNAGPRI 457

QY 493 CLGKDSAYLQMK-MALAILFRFYFRLLGHPVQYRMNTILSMAGHLKVRV 542

Db 458 CLGKDLAYLQMKSIASAVLLR-HRLTVTVGHKVEQKMSLTFMKYGLLVNV 507

RESULT 4

G86146

hypothetical protein F2214.14 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Dec-2001

C:Accession: G86146

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, C.W.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziali, R.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: G86146

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-554 <STO>

A:Cross-references: GB:AF005172; NID:g9920596; PIDN:AAF81318.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

C:Superfamily: Candida cytochrome P450 52A3; cytochrome P450 homology

Query Match 31.5%; Score 970; DB 2; Length 554;

Best Local Similarity 40.8%; Pred. No. 3.1e-67;

Matches 220; Conservative 86; Mismatches 169; Indels 44; Gaps 13;

QY 28 ALLLVVL---SWILVQWSLRKQKGRSPWVGATVQQLRNHHRMDLVGLY-----SRH 80

Db 6 AMLLVAIVTGYLWFKGIS-RWLKGRVWPLGLSLPLGIEQDRMHEWITENRACGGTY 64

QY 81 RTVVDMPFTS-----YTYIADPVNVEHLKTNFTNPKGIVYRSYMDVLLGDGIFNADG 135

Db 65 QTCFAPVPLAKKGQGLVTVTCDPKNLEHLKTRFDNPKGTQSGVSHDLGQIFNSDG 124

QY 136 ELWRKQRTASFEFASKNLRD-PSAIVEREYSLKLSGLSQASKAGKVVDVMDQELYMWTLL 194

Db 125 DTWLFQRTAALEFTRTLRQAMGRVNRGKILRFPCLPILATAQNAEPVDLQDLILRLTF 184

QY 195 DSIKVGVEIGTSLPDLSPENSFAQADPAANIITLRFIDP--LWRKRFHHVGSALL 252

Db 185 DNICLAFGKQTRTCAPGLPENGFAFADRAEASLQRFIIPKFMWKLKWLGLGLEVSL 244

QY 253 AOSIKLVDEFYTVSIVRRKAEIVRASKGQKMKHDLTSLRFIELGEAGDDGGGFGDDKS 312

Db 245 SRSGEIDEYLAUVINTRKQLMSQESGTHOR-HDDLILSRPMKKTES-----YSTTF 297

QY 313 LRDVVLNFIAGROTTATTLTWSFTTHAMSHPDVAEKLRLRELCAPAEARAREEGVTLVLCG 372

Db 298 LQHVALLILAGDRTSSVALSWFFLIITMHPTEVDKIVREICS-----VLIE 344

QY 373 GADADDKFAARVAQFAGLLTYDSLGLKLYLHACVTEITLRLYPVAPQDPKGLLEDVLPD 432

Db 345 TRGTDDVASWTSEP-----LGDELDRLVYLKAAISEITLRLYSPVDSKXHVENDVLPD 399

QY 433 GTRKVRAGGVTVPYSGRMENWGPDAASFRPERWINE-DGAFRNASPFKFTAFQAGPR 491

Db 400 GTFVPAGSVTYSIYAAGRMKSTWGEDCLEFPERWISPIDGKFINHDQYRFVAFNAGPR 459

QY 492 ICLGKDSAYLQMK-MALAILFRFYFRLLGHPVQYRMNTILSMAGHLKVRVSRVCHG 549

Db 460 ICLGKDLAYLQMKTTAAAVLLR-HRLTVTVGHKVEQKMSLTFMKNGLLVNLKRDLOQ 517

RESULT 5

B96662

probable cytochrome P450 F24D7.10 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001

C:Accession: B96662

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziali, R.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: B96662

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-523 <STO>

A:Cross-references: GB:AF005173; NID:g6456162; PIDN:AAF09150.1; GSPDB:GN00141

C:Genetics:

A:Gene: F24D7.10

A:Map position: 1

C:Superfamily: Candida cytochrome P450 52A3; cytochrome P450 homology

Query Match 31.5%; Score 969.5; DB 2; Length 523;

Best Local Similarity 39.7%; Pred. No. 3.1e-67;

Matches 211; Conservative 99; Mismatches 173; Indels 49; Gaps 12;

QY 27 IALLVVLWILVQWSLRKQKGRSPWVGATVQQLRNHHRMDLVGLY-----SRHRT 82

Db 10 LTLIVTVIIFVSLR---RSYKGRVWPLVGLSLPALITNAHRMHDFTADNLMCGGTYQT 66

QY 83 VTVDMPFTS-----YTYIADPVNVEHLKTNFTNPKGIVYRSYMDVLLGDGIFNADGEL 137

Db 67 CTFPPLAKKGQGLVTVTCDPKNLEHLKTRFDNPKGTQSGVSHDLGQIFNSDGT 126

QY 138 WRKQRTASFEFASKNLRDPSA-IVFREYSLKLSGLSQASKAGKVVDVMDQELYMWTLLS 196

Db 127 WRFQRTAALEFTRTLRQAMGRVNRGKILRFPILPILATESARABPIDLQDVLRLTFDN 186

QY 197 ICKVGVEIGTSLPDLSPENSEAQADPAANIITLRFIDP--LWRKRFHHVGSALLA 254

Db 187 ICGTLFGKQRTSLPPEPFENGFAVAFQATQATLQRIIMPEFIWKIRKWLRLGLEDDMSR 246

QY 255 SIKLVDEFYTVSIVRRKAEIVRASKGQKMKH-DILSRFIELGEAGDDGGGFGDDKSL 313

Db 247 SISHVDNYLSEIINTRKLELL---GQODESRHDDLILSRFMKKES-----YDKYL 295

QY 314 RDVVLNFIAGROTTATTLTWSFTTHAMSHPDVAEKLRLRELCAPAEARAREEGVTLVLCG 373

Db 296 KYVALNFIILAGDRTSSVALSWFFFWLVSINPRVEEKIINEIC-----TIL--- 340

QY 374 ADADKFAARVAQFAG-LLTYDSLGLKLYLHACVTEITLRLYPVAPQDPKGLLEDVLPD 432

Db 341 -----KTRDNTNKSMTDEPLTFDEIDQLVYLKAAISEITLRLYSPVDSKXHVENDVLPD 395

QY 433 GTRKVRAGGVTVPYSGRMENWGPDAASFRPERWINE-DGAFRNASPFKFTAFQAGPRI 492

Db 396 GTFVPSGNTVTSYISVGRMKFIWGEDCLEPKPERWL--EESRDEKQNYKFVAFNAGPRI 454  
Qy 493 CLGKDSAYLOMKMALAILFRYSFRLLGHPVQVVRMTILSMAGHLKVRYSR 544  
Db 455 CLGKDLAYLOMKSTASILLRHRUTVAPGHRVQRQMSLTLEFMKFLGKMDVHK 506

RESULT 6  
T00404  
Probable cytochrome P450 At2g44890 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 02-Feb-2001  
C:Accession: T00404; A84884  
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Masc  
submitted to the EMBL Data Library, July 1997  
A:Description: Arabidopsis thaliana chromosome II BAC Tl3E15 genomic sequence.  
A:Reference number: Z14146  
A:Accession: T00404  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-490 <ROU>  
A:Cross-references: EMBL:AC002398; NID:G3420042; PID:G2344895  
A:Experimental source: cultivar Columbia  
A:Map position: 2  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-769, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUIB:20083487; PMID:10617197  
A:Accession: A84884  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-490 <STO>  
A:Cross-references: GB:AE002093; NID:G2344895; PID:AA031835.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g44890; Tl3E15.10  
A:Map position: 2  
A:Introns: 33/2; 153/3; 361/3; 428/3  
C:Superfamily: Candida cytochrome P450 52A3; cytochrome P450 homology  
F:281-456/Domain: cytochrome P450 homology <P45>

Query Match 29.5%; Score 907.5; DB 2; Length 490;  
Best Local Similarity 40.7%; Pred. No. 1.8e-62;  
Matches 193; Conservative 85; Mismatches 167; Indels 29; Gaps 6;  
Qy 67 HEMHDLVGYLSRHRVTVDMPFTSYTADPNVVEHLKTNFTNPKGIVVSYMDVLL 126  
Db 34 HLYDYETAIATKTFTRFLSPGQSEIFTADPRNVEHLKTRFHNYSKGPVGTNADLL 93  
Qy 127 GDGIFNADGELWRKQKRTASPFASKNLRDFAIVFREYSLSKLSGILSQASKAGKVDVQ 186  
Db 94 GHGIFAVDGEKWKQKLVSPFSTRVLNFSYVFTSASKLVGFIAEFALSGKSPDFQ 153  
Qy 187 ELYMRTLDISKGVGVEIGTISPDLEP--NSFAQAFDAANIITLRIIDPLWRKRFH 245  
Db 154 DMLMCKTLDISFKVGFVGLGDLGFSKEGEBFMKAFDEGNAGTSRVTDPFWKLCFLN 213  
Qy 246 VGEALLAQSIKLVDFTYSVTRRRKAEIVFVRASGQKQKMHDLTSLRFLGEAGDDGG 305  
Db 214 IGSERLKKSIALLDKFVYSLITTRKEL----SKEQNTSVREDLSKFLLESEKDPEN- 268  
Qy 306 GFDDKSLRDVNLNFIAGRTTATLSWFTTHAMSHPDVAEKRLRELCAPFAERAREEG 365  
Db 269 --MNDKYLRIILNVVAGKDTTASLSWFLVLCNPLVQEKIVQETRDVTSSEKTD 326  
Qy 366 VTILVCGGADADKAPAAVAQFAGLLTYDLSGLKLVYLHACVTEILRYLPVAPQPKGL 425  
Db 327 VN-----GFIESVTEALAQOMQLHAALSETMRLYPPVPEMRCAE 367  
Qy 426 EDDVLPGTKVRAGMVTVPVYSGMGMENWGPDAASPRPERWINEGDAFRNASPFKFTA 485  
Db 368 NDDVLPGHRVSKGNDIYYISVAMGRWTVINGQDAEEFKPERWL--KGVFQESQFKFIS 426

Qy 486 FORAPRICGKDSAYLOMKMALAILFRYSFRLL--EGHPVOYVRMTILSMAGHL 538  
Db 427 FHAGPRICIGKDFAYRQMKIVSMALLHFFRFRKADENSKVSYKKMLTLTHVDGGL 480  
RESULT 7  
T50510  
cytochrome P450-like protein - Arabidopsis thaliana  
N:Alternate names: protein T22D6.190  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 16-Feb-2001  
C:Accession: T50510  
R:Bevan, M.; Terry, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R.; De  
ewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.  
submitted to the Protein Sequence Database, May 2000  
A:Reference number: Z25101  
A:Accession: T50510  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-550 <BEV>  
A:Cross-references: EMBL:AL357612  
A:Experimental source: cultivar Columbia; BAC clone T22D6  
C:Genetics:  
A:Map position: 5  
A:Introns: 417/3  
A:Note: T22D6.190  
C:Superfamily: Candida cytochrome P450 52A3; cytochrome P450 homology  
F:330-508/Domain: cytochrome P450 homology <P45>

Query Match 29.4%; Score 904.5; DB 2; Length 550;  
Best Local Similarity 39.4%; Pred. No. 3.7e-62;  
Matches 212; Conservative 97; Mismatches 182; Indels 47; Gaps 17;  
Qy 29 LLLVVLGMLVQVRSRKQK--GPRSWFVIQ---ATVEQLRNRYHRMHDWLGVYL--SRHRT 82  
Db 40 LELFIAFFVATIHSLRQKKHQGPVWFLVGLMFLSLISAVRS--NIVELSDVLISQNGT 97  
Qy 83 VTVDMPFTS---YTYIADPNVVEHLKTNFTNPKGIVVSYMDVLLGDGIFNADGELWR 139  
Db 98 PRFRGPFSTLNCVTCDPNVEHLKTRFSIYKGSYFRETMDLLGDLGIFNTDDGTWQ 157  
Qy 140 KQRTASPEFASKNLRDFAIVFREYSLSKLSGILSQASKAGKVDVMDLYMRTLDISCK 199  
Db 158 RQKAAASVEFHSAPKQLTQSLSLHE--LVHNRLLPVLETSGK--IDLQILLRLTFDNCVM 214  
Qy 200 VGFVGEIGTSLPDLSPENSFAQAFDAANIITLRIIDP--LWRIKRFPHVGSALLAQSIK 257  
Db 215 IAFGVDPCLSPKLPPIPAKAFADATEATVVRVMPKFWKLMRSNLNLTGKKLKESIN 274  
Qy 258 LVDBFTYSVTRRRKAEIVFVRASGQKQKMH--DILSRFIELGEAGDGGGFGDDKSLRDV 316  
Db 275 GVDDFAFEBVIRTKEM-----SLETEIAKFPDLITIFMGL---RDENGQKFSKFLRDI 326  
Qy 317 VLNFIAGRTTATLSWFTTHAMSHPDVAEKRLRELCAPFAERAREEGVTILVCGGADA 376  
Db 327 CVNFIAGRTTSSVALSNFFWLIENKPEVEBKIMWGCKILEQV----- 371  
Qy 377 DDKAFARVAQFAGLLTYDLSGLKLVYLHACVTEILRYLPVAPQPKGILEDDVLPGTKV 436  
Db 372 --DHGDTKKNMEYEFVFRPEEIKKMDYLQAALSETLRLYPSVFDHKEVLEDDVDPDGTKL 430  
Qy 437 RAGMVTVPVYSGMGMENWGPDAASPRPERWINEGDAFRNASPFKFTAPQAGRICLGK 496  
Db 431 KKGSKVIYAIYAMGRMETINGKDCREKPERWL--RDGRYMSSESAIKFTAFNGGPRCLGK 489  
Qy 497 DSAYLOMK--MALAILFRYSFRILLE--CHPVOYVRMTILSMAGHLKVR--VSRVCHGD 550  
Db 490 DFAYQMRVAAAIYR--YKVRVDKGGHKVPEPKMALTNYMKHGLKVNVRVRSSEID 546

RESULT 8  
T49064

cytochrome P450-like protein - Arabidopsis thaliana  
 N:Alternate names: protein TSP19.280  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 16-Feb-2001  
 C:Accession: T49064  
 R:Benes, V.; Wurmbach, E.; Drzonek, H.; Ansoorek, W.; Mewes, H.W.; Rudd, S.; Lemcke, K.; submitted to the Protein Sequence Database, April 2000  
 A:Reference number: Z25014  
 A:Accession: T49064  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-499 <BEN>  
 A:CROSS-references: EMBL:AL163972; GSPDB:GN00061; ATSP:TSP19.280  
 A:Experimental source: cultivar Columbia; BAC clone TSP19  
 C:Genetics:  
 A:Gene: ATSP-TSP19.280  
 A:Map position: 3  
 A:Insertions: 153/2  
 C:Superfamily: Candida cytochrome P450 homology  
 F:295-467/Domain: cytochrome P450 homology <P45>

Query Match 28.6%; Score 879.5; DB 2; Length 499;  
 Best Local Similarity 38.8%; Pred. No. 2.8e-60;  
 Matches 205; Conservative 96; Mismatches 176; Indels 51; Gaps 13;

QY 27 IALL--VLSWLVORWSLRKQKGRSPVIGATVEQLRNHYHRMDLWLVGYLSRHRTVT 84  
 DB 7 ISLLCPVIVFFPPKNS--SSEFGKSPVIGSLPGLVNNRRFLDWTVETLSRCPTQT 65  
 QY 85 VDM---PFTSYIADPNVNEHLKNTFTPKYKIVRSYMDVLLGGIINAGELWRKQ 141  
 DB 66 AIFRRPKLQFVMTANPANVEYMLKTFSPFKGERFISILEDFLRGIFNSDGMWKKQ 125  
 QY 142 RKTASPFASKNLDF--SAIVREYSKLSGILSOASKAGKVDMDQSLYMRMTLSDICKV 200  
 DB 126 RKTASVEFTSKURDFVMSNVTVETNRLVPLVLAETAKNGLDQLDLRPFADNICKL 185  
 QY 201 GFGVEIGTLPDLPEN--SFAQAFDAANIIITLRF---IDPLWRKRPFFHVSSEALLAQSI 256  
 DB 186 AFNVDSACLGDGAGVNFQAFETAATIIISQRFQSVISYKWKIKKLNIGSERVLRSEI 245  
 QY 257 KLVDEFTSYVIRRKAEIVREVRASGKQKMDILSRFIELGEAGDGGGFGDDKSLRDV 316  
 DB 246 MIVHKFADEIVRNR---IE---QGVSDSHKEDLLSRFISKEE-----MNSPEILRDI 291  
 QY 317 VLPNVIAGRTTATTLSTFWTHMSPDVAEKLRRLECAFEARERBEGVTLVLCGGADA 376  
 DB 292 VISFILAGRTTSSALLSNFFWLLSMHPEVKDKILQELNSIRPTGK----- 337  
 QY 377 DDKAFAARVAQFAGLLTYDSLGKLVYLHACVTETRLYPVAPQPKGILEDDVLPDGTKV 436  
 DB 338 -----RIGEVYG---PEDLKMNYLHAAITSLRLYPPVVDVTMSCAEDNVLPDGTFI 387  
 QY 437 RAGMWVTVYPSYSGRMENYNGWPDAAFRPERWINE--DCAFRNASPPFKETAQAGPRICLG 495  
 DB 388 GKDWGISYNAVANGMESINGKDCDRFDPERWIDETNGFGFNGENPKYKPFAPFAGERMCLG 447  
 QY 496 KDSAYLQMK--MALAILFRYSRLLEGHVPQVYRMVMTILSMAGHLKVRV 542  
 DB 448 KEMAYIQMKSIVAALIERFVVEVPCKRPERPEILMSVTLIRIGGLNVRV 495

RESULT 9  
 JC5965  
 Cytochrome P450 CYP86A1 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 28-Jul-2000  
 C:Accession: JC5965  
 R:Benveniste, I.; Tijer, N.; Adas, F.; Philipps, G.; Salauen, J.P.; Durst, F.; Biochem. Biophys. Res. Commun. 243, 688-693, 1998  
 A:Title: CYP86A1 from Arabidopsis thaliana encodes a cytochrome P450-dependent fatty aci  
 A:Reference number: JC5965; MUID:98162595; PMID:9500987  
 A:Accession: JC5965

A>Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-513 <BEN>  
 A:CROSS-references: EMBL:X90458  
 C:Superfamily: Candida cytochrome P450 homology  
 F:300-478/Domain: cytochrome P450 homology <P45>

Query Match 28.5%; Score 876.5; DB 2; Length 513;  
 Best Local Similarity 39.6%; Pred. No. 5e-60;  
 Matches 210; Conservative 83; Mismatches 182; Indels 55; Gaps 13;

QY 27 IALLLVLSWLVORWSLRKQKGRSPVIGATVEQLRNHYHRMDLWLVGYL-----SRHRT 82  
 DB 13 VAAISVYALWFY---FLSRLLTGPKVLPVGSLLPYLIANPSRIHDWIADNLRTATGTYQT 69  
 QY 83 VTVDMPFTS-----YTYIADPNVNEHLKNTFTPKYKIVRSYMDVLLGGIINAGEL 137  
 DB 70 CTWVTFVAKAQQFYVTTCPKXNVEHLKTRFDNTPFKGPMWRAAFHDLGQGFNSDGT 129  
 QY 138 WRKQRTASPFASKNLDRFSA--IVFREYSKLSGILSOASKAGKVDMDQSLYMRMTLDS 196  
 DB 130 WLMQRKTAALFETTRTLROAMRWVNGTINKRLWLLDRVQNNKFPVDLQDLFLRLTFDN 189  
 QY 197 ICKVGFGEIGTLPDLPEN--SFAQAFDAAN--IITLRFIDPLWRKRPFFHVSSEALLAQ 254  
 DB 190 ICLTFGKDPQLSLDLPDPNPSVAFDTATEATLRELLYTGFLWRIQKAMGIGSEDKLKK 249  
 QY 255 SKLVDEFTSYVIRRKAEIVREVRASGKQKMDILSRFIELGEAGDGGGFGDDKSLR 314  
 DB 250 SLEVVETYNDAIDARK-----NSPDDLLSRFL-----KKRDVNGVNLPTDVLQ 294  
 QY 315 DVVLNVIAGRTTATTLSTFWTHMSPDVAEKLRRLECAFEARERBEGVTLVLCGGA 374  
 DB 295 RIALNFVLGRDTSSVALLSWFFFWLVMNNREVEYKINEL-----SMVLKEIR 341  
 QY 375 DADKFAARVAQFAGLLTYDSLGKLVYLHACVTETRLYPVAPQPKGILEDDVLPDGT 434  
 DB 342 GNDQEKWTEPLEF-----DEADRLVYLKAAAEATRLYPSVPQDFKYVVEHDLVDPGT 395  
 QY 435 KVRAGMWTVYPSYSGRMENYNGWPDAAFRPERWINE--DCAFRNASP---FKETAQAGPR 491  
 DB 396 FVPRGSTVTVYSYSGRMKTINGEDCLFPRPRLMTADGE--RFETPKDGYKFFVAFNAGR 454  
 QY 492 ICLGKSAYLQMK--MALAILFRYSRLLEGHVPQVYRMVMTILSMAGHLKV 540  
 DB 455 TCLGKDLAYNQMKSVASAVLLIRYRFP--VEGHRVEQKMSLTLFKNGLRV 503

RESULT 10  
 B86379  
 protein F21J9.20 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C:Accession: B86379  
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C. C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719; PMID:11130712  
 A:Accession: B86379  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-522 <STO>  
 A:CROSS-references: GB:AE005172; NID:g9743340; PIDN:AAF97964.1; GSPDB:GN00141  
 C:Genetics:  
 A:Gene: F21J9.20  
 A:Map position: 1

C;Superfamily: Candida cytochrome P450 52A3; cytochrome P450 homology

Query Match 27.2%; Score 837; DB 2; Length 522;

Best Local Similarity 36.2%; Pred. No. 6e-57;

Matches 192; Conservative 83; Mismatches 204; Indels 52; Gaps 11;

QY 27 IALLVLSWILVQWLSRQKQPRSPVIGATVEQLRNYHRMHDMLV-----GYLSRH 80  
DB 20 VALALVGLFLLSYLREKLVSKGPPVMPVLGIIPMLALNKHDLFTWCTRCVRSRGTFHY 79  
QY 81 RTVTVDMPFTSYTYIADPNVVEHVILKTNFTNPKGIVYRSMYMDVLLGDGIFNADGELWR 140  
DB 80 RGIWFGAYGIMT--ADPANVEHILKTNFTNPKGIVYRSMYMDVLLGDGIFNADGELWR 137  
QY 141 ORKTASFEFASKNLRDFAIVRE-YSLKLSGILSQAASKGVKVDMDQELVYRMFLDSICK 199  
DB 138 ERRVAKTEHMSRPLEHTFTMRDLVDQKLVPLMENLSTSKRVFEDLQDLRLFRFTDNICI 197  
QY 200 VGFGEIGTSLPDLSPENSFAQADAAANIITLRFIDP--LMRIKRFHVHVSSEALLAOSIK 257  
DB 198 SAFGVYPSLETGLPEIPFAKAFEDATRYTLARFLIPFVWPKMRFILGIGYERKLNNVR 257  
QY 258 LVDEFTSVYRRAKAEIVVEVRSQKQKMH-----DILSRFIEL---GEAGDDGGGFGD 309  
DB 258 IVHAFANKTVRERR-----NMRKLGINDYADLLSLMOREVEYKESDITTRGNYFS 308  
QY 310 DKSIRDVLNFIAGRDTTATTLGWFTHMAMSHPDVAEKLREELCAFEAREAREEGVTLV 369  
DB 309 DKYREFCTSIAGRDTTVALVWFVWLVQKHVEVEXRIILREIKRKLITQE----- 363  
QY 370 LCGGADADKAFARVAQFAGLLTYDSLGKLVYLHACVETELRYLPAVPQDPKGLLEDV 429  
DB 364 -----TED-----EDFRENVLQAALTESRLYPSVPMEMKQALEDDV 405  
QY 430 LPDGTQVAGGVTVYVPSYSGRMENWGPDAASFPPEMINEDGAFRNASPFKFTAFQAG 489  
DB 406 LPDGTQVAGGVTVYVPSYSGRMENWGPDAASFPPEMINEDGAFRNASPFKFTAFQAG 464  
QY 490 PRICLGKDSAYLQMKWALAILFRFYSFRLLGHPVQYRMMTILSMAHGLKV 540  
DB 465 PRLCVGKFAFYQMKWAAAILMRYSKVQVQCEIVPKLITLYLMKMGNTV 515  
RESULT 11  
T46196  
Cytochrome P450-like protein - Arabidopsis thaliana  
N:Alternate names: protein T8P19.30  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 16-Feb-2001  
C:Accession: T46196  
R:Choi, N.; Robert, C.; Brattier, P.; Wincker, P.; Cattolico, L.; Artiguenave, F.; Sa  
submitted to the Protein Sequence Database, December 1999  
A:Reference number: 223008  
A:Accession: T46196  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-506 <CHO>  
A:Cross-references: EMBL:AL133315  
A:Experimental source: cultivar Columbia; BAC clone T8P19  
C:Superfamily: Candida cytochrome P450 52A3; cytochrome P450 homology  
A:Map position: 3  
A:Note: T8P19.30  
Query Match 26.7%; Score 823; DB 2; Length 506;  
Best Local Similarity 36.1%; Pred. No. 6.9e-56;  
Matches 193; Conservative 92; Mismatches 188; Indels 62; Gaps 14;  
QY 27 IALLVLSWILVQWLSRQKQK-----GPRSPVIGATVEQLRNYHRMHDMLVGY--LSRH 80  
DB 9 ILAFLITIFLSSSSSTKQVQENTYGPSPYPLIGSILSPKNRHLLQWYELLRLSPS 69

QY 81 RTVTVDMPFTSYTYIADPNVVEHVILKTNFTNPKGIVYRSMYMDVLLGDGIFNADGELWR 139  
DB 69 QUILVPELLGNRRITITNPNVEYILKTNFFNPKGKFTDILGLLGGGIFNVDGHSNS 128  
QY 140 KOKKTASFEFASKNLRDFAIVRE-YSLKLSGILSQAASKGVKVDMDQELVYRMFLDSIC 198  
DB 129 SQKSLASHSEFSTRSLSFAFVLKQBEVENLVPVLSAADVGTTVDLQDLVLRFAFDVVC 188  
QY 199 KVGFGVEIGTSLPDLSPENSFAQADAAANIITLRFIDP--LMRIKRFHVHVSSEALLAOS 255  
DB 189 KVSGLWDPDCLDLTRVNPVLEAFDAEISARRATEPIYAVMKTTRVLNVGSEKRLRA 248  
QY 256 IKLVDEFTSVYRRAKAEIVVEVRSQKQKMHDIILSRFIELCEAGDDGGGFGDGLSLRD 315  
DB 249 IRVTHVILVSEIVRAKKSL-EI---GTGAQKQDILSRFLAAGHGE-----AVRD 295  
QY 316 VLNFVFIAGRDTTATTLGWFTHMAMSHPDVAEKLREELCAFEAREAREEGVTLVLCGGAD 375  
DB 296 MVISFINAGRDTTSAAMTWLFWLLTENDDVERKI-----LEEVDPLVSLG--- 340  
QY 376 ADDKAFARVAQFAGLLTYDSLGKLVYLHACVETELRYLPAVPQDPKGLLEDVLPDGT 435  
DB 341 -----LGPEDEKEMAYTKACLEAMRLYPPVSWDSKHAANDDVLDPDGT 384  
QY 436 VRAGGMVTVYVPSYSGRMENWGPDAASFPPEMINEDGAFRNASPFKFTAFQAG 490  
DB 385 VKRGDKVTVYVPSYSGRMENWGPDAASFPPEMINEDGAFRNASPFKFTAFQAG 444  
QY 491 RICLGKDSAYLQMKWALAILFRFYSFRLLGHPVQYRMMTILSMAHGLKVRSR 544  
DB 445 RVCVGKEMAFQMKVYVGSVLSRFEIVPVNKRDPVFPVPLLT-AHMAAGLKVVKIR 498  
RESULT 12  
G86265  
F3P19.17 protein - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 27-Nov-2001  
C:Accession: G86265  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A96141; MUID:21016719; PMID:11130712  
A:Accession: G86265  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-529 <STO>  
A:Cross-references: GB:AE005172; NID:G4850397; PIDN:AD31067.1; GSPDB:GN00141  
C:Genetics:  
A:Map position: 1  
C:Superfamily: Candida cytochrome P450 52A3; cytochrome P450 homology  
Query Match 25.6%; Score 788; DB 2; Length 529;  
Best Local Similarity 34.2%; Pred. No. 3.8e-53;  
Matches 187; Conservative 102; Mismatches 193; Indels 64; Gaps 13;  
QY 24 HKYIALLVLSWILVQWLSRQKQKPRSPVIGATVEQLRNYHRMHDMLVGLS 78  
DB 8 YNHLISLFDLLSLGLFVFCCLREKLTNRKGFMLPVFGITLFFHINDVYVWTRSLK 67  
QY 79 R-----HRTVDMPTSYTYIADPNVVEHVILKTNFTNPKGIVYRSMYMDVLLGDGIFN 132  
DB 68 KCRGFLYNGWLDGSGYAVTCV--PANVEYMLKTNFKNFPAKTFKSRFNDLLEGIEN 125  
QY 133 ADGELWRKORTASFEFASKNLRDFAIVREYSL-----KLSGILSQAASKGV 183

Db 126 ADDSWKQRRLLIITEMHSTG-----FVHSGFQTTQHLVRKLLKVMESFAKSOBAF 177  
 QY 184 DMQELYMRTLDSDICKVGFVEIGTSLSPDIPNSFAQAADFANIIITLRFIDP--LWRIK 241  
 Db 178 DLQDVFLRLTFTDIIICLAGLGADPETLAVDLFPQPPAKAFEAATESLTFRRFPIPPKPM 237  
 QY 242 RPFHVGSEALLAQSIKLVDFTYSVIRRKABIVVRASGKQEK--KHDILSRFIEL-G 298  
 Db 238 RLDTGYEKGLRIAGVGVHGFVDKMIVDRIICELKE-----EETLDRSDVLRIQIES 291  
 QY 299 EAGDGGGFGDDKSLRDVVLNFIAGRTTATTLNFWTHMAMSHPDVAEKLRELCAE-E 357  
 Db 292 HKRENEIDPSTIRFRQFCTSPILAGRTSSVALSWFCWVIQKHEVENKI---ICEIRE 348  
 QY 358 ABEAREEGTVLVLGGADADDKFAARVAQFAGLLTYDSLGLKVLVHACVTELTLLYPAV 417  
 Db 349 ILRQGRDSTP-----SKNESLFTVVKELNMMVLOALSETLFLFPI 390  
 QY 418 PDGPKGILEDDVLPDGTCKVRAGMVTYVPSYGRMEYNWGPDAASFRPRERWINEGAFRN 477  
 Db 391 PHEMKCAIEDDVLDPDGTFRKGRSVVFSIYANGRMESIWGKDCEIFRPRERWI-QAGKFS 449  
 QY 478 ASPFKETAFQAGPRICLGDSDAYLQMKMALILFRYSFRLLECHPVQVRMTILSMANG 537  
 Db 450 DDQFKVWFENAGPRLCIGKTFAYLQMKMIAASVLLRYSIKVQDQHVIAPRVTNLYMYKG 509  
 QY 538 LKVRVS 543  
 Db 510 LKVTIT 515  
 RESULT 13  
 T08014  
 cytochrome P450 CYP94A1 - spring vetch  
 C:Species: *Vicia sativa* (spring vetch, tare)  
 C:Date: 21-May-1999 #sequence\_revision 21-May-1999 #text\_change 16-Feb-2001  
 C:Accession: T08014  
 R:Tijer, N.; Helwig, C.; Pinot, F.; le Bouquin, R.; Lesot, A.; Durst, F.; Salaun, J.P.;  
 Biochem. J. 332, 583-589, 1998  
 A:Title: Functional expression in yeast and characterization of a clofibrate-inducible P  
 A:Reference number: Z16287; MUID:98264856; PMID:9601090  
 A:Accession: T08014  
 A:Status: preliminary;  
 A:Molecule type: mRNA  
 A:Residues: 1-514 <T1J>  
 A:Cross-references: EMBL:AF030260; NID:g4204094; PIDN:AAD10204.1; PID:g4204095  
 A:Gene: vagn11  
 C:Superfamily: Candida cytochrome P450 52A3; cytochrome P450 homology  
 F:312-480/Domain: cytochrome P450 homology <P45>  
 Query Match 25.5%; Score 785.5; DB 2; Length 514;  
 Best Local Similarity 35.0%; Pred. No. 5.8e-53;  
 Matches 165; Conservative 95; Mismatches 154; Indels 57; Gaps 8;  
 QY 50 PRSPEVIGATVEQURNYHRMMDL--VGYLSRHRTVTVDMPFTSYTI-ADPVNVEHLVK 106  
 Db 48 PKSYPLIGSYLSPKRLHRRQLWLSDIVQISPSATFQDGLGKRLITGNPSVQHIK 107  
 QY 107 TNFTNPKGIVYRMYDVLGDTFNADGELWRKQRTASPEFASKNLRDF-SALVREY 165  
 Db 108 NQFSNYQKGTFTNTSLDFLTGTFNTNGPNWKFQVQASHEFNKTRIRNFEVHIDTEL 167  
 QY 166 SLKLSGLSQSKAGKGVDMQELYMRWTLDSICGVGVEIGTISPOLPENSFAQFADAA 225  
 Db 168 TNRILPILTSQTNWILDFQDILQRFTFDNICMIAFGYDEYITPSTNRKFAEVEDA 227  
 QY 226 NIITLRFIDPL--WRTRKFRHVGSEALLAQSIKLVDFTYSVIRRKABIVVRASGK 282  
 Db 228 TEISKRRLFLPLPIWIKKIVFNIGSEKELAEVTEVRSFAKLVRKRELEB-----K 282  
 QY 283 QEKMKHDLRFIELGEAGDGGGFGDDKSLRDVVLNFIAGRTTATTLNFWTHMAMSH 342

Db 283 SSLETDMLSRFL-----SSCHSDEDFVADIVISIFILAGKDTTSAALTWFELLWKN 334  
 QY 343 PDVAEKLRELCAFEAREABREBEGTVLVLGGADADDKFAARVAQFAGLLTYDSLGLVY 402  
 Db 335 PRVEEETVNEU-----SKGBELMAYDEVKEMVY 362  
 QY 403 LHACVTELTLLYPAVPQDPKGILEDVLPDGTCKVRAGMVTYVPSYGRMEYNWGPDAAS 462  
 Db 363 THAALSMSRLYPPVPMDSKAEAVNDVLPDGVVKKGTIVTYHYVAMGRMKSILGDDWAE 422  
 QY 463 FRPERWINE-----GAFRNASPPFTAFQAGPRICLGDSDAYLQMKMALA 508  
 Db 423 FRPERWLEKDEWNGKWFVGRDSYSPVQFAGPRVCLGKEMAFQMKRIVA 473  
 RESULT 14  
 F86265  
 hypothetical protein F3F19.16 - Arabidopsis thaliana  
 C:Species: *Arabidopsis thaliana* (mouse-ear cress)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Dec-2001  
 C:Accession: F86265  
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
 ausen, N.P.; Hughes, B.; Huizar, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
 Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.  
 A:Reference number: A86141; MUID:21016719; PMID:11130712  
 A:Accession: F86265  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-519 <STO>  
 A:Cross-references: GB:AE005172; NID:g4850398; PIDN:AAD31068.1; GSPDB:GN00141  
 C:Genetics:  
 C:Superfamily: Candida cytochrome P450 52A3; cytochrome P450 homology  
 Query Match 25.3%; Score 780; DB 2; Length 519;  
 Best Local Similarity 34.1%; Pred. No. 1.6e-52;  
 Matches 184; Conservative 91; Mismatches 187; Indels 78; Gaps 12;  
 QY 27 IALLVVLVSWILVQWLSLRKQKQPRSPVIGATVEQLRNHYHRMHDMLVGLSR-----H 80  
 Db 23 VSLALLGLFVCCVREKVKLGTITWVPFGITPEFFHRNDVYGVATRCCKRCRGTFLY 82  
 QY 81 RTVVDMPFTSYTIADPVNVEHLVKTNTNPKGIVYRMYDVLGDTGGINADGELWRK 140  
 Db 83 NGVWLGSGYGVATCV--PANVEYMLKTNFNKFPKGAFFKFRFNDLEDGIFNADAESWKE 140  
 QY 141 ORKTASPEFASKNLRDFAIVFREYSL-----KLGLSQSKAGKGVDMQELYMR 191  
 Db 141 ORLIITEMHSTR-----FVHSGFQTTQHLVRKLLKVMESFAKSOBAF 192  
 QY 192 MTLDSICKVGFVEIGTSLSPDIPNSFAQAADFANIIITLRFIDP--LWRIKFEHVGSE 249  
 Db 193 LTFDNICIAGLDGDPGTLSDPLVPPAQAFAFEATETMTFRFMPPIPKLFFFDIGYE 252  
 QY 250 ALLAQSTKLVDFTYSVIRRKABIVVRASGKQEKMHDLISRFTELGEAGDGGGFGD 309  
 Db 253 KGLRKAV-----DVSMSLSLSTR---WLIVVSASSKKKFSQSHKT-----D 288  
 QY 310 DKS-----LRDVVLNFIAGRTTATTLNFWTHMAMSHPDVAEKLRELCAFEARE 363  
 Db 289 EKDPSTIKFRQFCTSPILAGRTSSVALTWFFVVIQKHEVENKIIREI-----SEILRQ 344  
 QY 364 EGTVLVLGGADADDKFAARVAQFAGLLTYDSLGLKVLVHACVTELTLLYPVQDPKG 423  
 Db 345 RG-----DSPTSKNESLFTVVKELNMMVLOALSETLFLFPIPMEMKQ 388

Search completed: March 2, 2004, 15:29:43  
Job time : 22 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 7, 2004, 15:20:03 ; Search time 900 Seconds  
(without alignments)

15836.681 Million cell updates/sec

Title: US-10-021-657-7

Perfect score: 3897

Sequence: 1 gaattccaagcaggccctt.....gctgagtaaacgacgaattc 3897

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2421054 seqs, 182871629 residues

Total number of hits satisfying chosen parameters: 4842108

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*

- 1: /cgn2\_6/ptodata/1/pubna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubna/US06\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubna/US06\_PUBCOMB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubna/US07\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubna/US09A\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubna/US09B\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/1/pubna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/1/pubna/US09D\_PUBCOMB.seq.\*
- 13: /cgn2\_6/ptodata/1/pubna/US10A\_PUBCOMB.seq.\*
- 14: /cgn2\_6/ptodata/1/pubna/US10B\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/1/pubna/US10C\_PUBCOMB.seq.\*
- 16: /cgn2\_6/ptodata/1/pubna/US10\_NEW\_PUB.seq.\*
- 17: /cgn2\_6/ptodata/1/pubna/US60\_NEW\_PUB.seq.\*
- 18: /cgn2\_6/ptodata/1/pubna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3897	100.0	3897	13	US-10-021-657-7
2	3897	100.0	3897	14	US-10-412-000-7
3	1378.4	35.4	1306	13	US-10-021-657-1
4	1378.4	35.4	1306	14	US-10-412-000-1
5	1092	28.0	1092	13	US-10-021-657-5
6	1092	28.0	1092	14	US-10-412-000-5
7	268.4	6.9	494	13	US-10-021-657-3
8	268.4	6.9	494	14	US-10-412-000-3
9	263	6.7	267	13	US-10-021-657-6
10	263	6.7	267	14	US-10-412-000-6
11	164.2	4.2	1807	12	US-10-425-114-23429
12	154.2	4.0	825	12	US-10-425-114-35451
13	151.6	3.9	1617	15	US-10-260-238-1236
14	146.4	3.8	854	15	US-10-260-238-4619
15	143.6	3.7	1947	12	US-10-425-114-28465

16	142.8	3.7	1680	12	US-10-425-114-1524	Sequence 1524, Ap
17	138.4	3.6	554	15	US-10-260-238-4618	Sequence 4618, Ap
18	137.6	3.5	1227	15	US-10-260-238-422	Sequence 422, App
19	136.6	3.5	604	15	US-10-260-238-3255	Sequence 3255, Ap
20	132.4	3.4	444	15	US-10-260-238-5268	Sequence 5268, Ap
21	130	3.3	792	15	US-10-260-238-4608	Sequence 4608, Ap
22	128.2	3.3	1694	12	US-10-425-114-28590	Sequence 28590, A
23	127.8	3.3	658	15	US-10-260-238-4017	Sequence 4017, Ap
24	125.8	3.2	520	15	US-10-260-238-4018	Sequence 4018, Ap
25	124.4	3.2	1638	15	US-10-260-238-907	Sequence 907, App
26	118.4	3.0	1462	12	US-10-424-599-20189	Sequence 20189, A
27	117.2	3.0	1267	12	US-10-425-114-32933	Sequence 32933, A
28	115.8	3.0	620	12	US-10-425-114-4749	Sequence 4749, Ap
29	114.2	2.9	1806	12	US-10-425-114-28220	Sequence 28220, A
30	106.2	2.7	546	15	US-10-260-238-3287	Sequence 3287, Ap
31	105.4	2.7	1781	12	US-10-425-114-26392	Sequence 26392, A
32	104	2.7	1662	9	US-09-938-842A-881	Sequence 881, App
33	104	2.7	1662	11	US-09-938-842A-881	Sequence 881, App
34	98.4	2.5	1498	12	US-10-425-114-17454	Sequence 17454, A
35	98.4	2.5	2763	12	US-10-424-599-120026	Sequence 120026, A
36	96	2.5	1614	9	US-09-938-842A-2678	Sequence 2678, Ap
37	96	2.5	1614	11	US-09-938-842A-2678	Sequence 2678, Ap
38	95.6	2.5	351	15	US-10-260-238-5726	Sequence 5726, Ap
39	94.8	2.4	1729	12	US-10-425-114-880	Sequence 880, App
40	94.8	2.4	2431	12	US-10-424-599-120032	Sequence 120032, A
41	93.8	2.4	866	9	US-09-770-445-572	Sequence 572, App
42	93.2	2.4	2239	12	US-10-425-114-10296	Sequence 10296, A
43	92.4	2.4	1542	9	US-09-938-842A-1868	Sequence 1868, Ap
44	92.4	2.4	1542	11	US-09-938-842A-1868	Sequence 1868, Ap
45	90.2	2.3	800	12	US-10-424-599-58579	Sequence 58579, A

ALIGNMENTS

RESULT 1  
US-10-021-657-7  
; Sequence 7, Application US/10021657  
; Publication No. US20020083483A1  
; GENERAL INFORMATION:  
; APPLICANT: ALBERTSEN, MARC C.  
; APPLICANT: FOX, TIM  
; APPLICANT: HUFFMAN, GARY  
; APPLICANT: TRIMMELL, MARY  
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES MEDIATING MALE FERTILITY AND  
; TITLE OF INVENTION: METHOD OF USING SAME  
; FILE REFERENCE: 1148  
; CURRENT APPLICATION NUMBER: US/10/021,657  
; CURRENT FILING DATE: 2001-12-14  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 7  
; LENGTH: 3897  
; TYPE: DNA  
; ORGANISM: Zea mays  
US-10-021-657-7

Query Match	100.0%;	Score 3897;	DB 13;	Length 3897;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 3897;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
QY	1	GAATTC	CAAGCAGGCGCCCTTGTAGCAGAGAGTGTGTGCTGATCAGTTCGCGGAAATGAGT	60
Db	1	GAATTC	CAAGCAGGCGCCCTTGTAGCAGAGAGTGTGTGCTGATCAGTTCGCGGAAATGAGT	60
QY	61	CGGTGCTGAGAGCAACGCTGAGGGGTTCCAGGATGGCAATGGCTATGCGCAATCGGCTAG	120	
Db	61	CGGTGCTGAGAGCAACGCTGAGGGGTTCCAGGATGGCAATGGCTATGCGCAATCGGCTAG	120	
QY	121	AGGTGAGAGCAAGTGTGTGAGATTGGAGGCGACCTTATGCAAGTGTGTGAGAGGC	180	
Db	121	AGGTGAGAGCAAGTGTGTGAGATTGGAGGCGACCTTATGCAAGTGTGTGAGAGGC	180	

181	ACGCAATGAGAGATCTATTACAGACTTACACTGGATGCGGCCAACAAATTCACACCTTTAGA	240	QY
181	ACGCAATGAGAGATCTATTACAGACTTACACTGGATGCGGCCAACAAATTCACACCTTTAGA	240	Db
241	TTTTGATACTGTCACCTCTACTTTTATCTCTGGTTGGGCAACTCCATAGGCTCATGTT	300	QY
241	TTTTGATACTGTCACCTCTACTTTTATCTCTGGTTGGGCAACTCCATAGGCTCATGTT	300	Db
301	AATCAATGATTAGTGATATTACAGCAAAATATCTTGTTGGTTGGACATTTATAATATGTG	360	QY
301	AATCAATGATTAGTGATATTACAGCAAAATATCTTGTTGGTTGGACATTTATAATATGTG	360	Db
361	GGGTGAGACGGATTAAATATCATCCATGAGAGCTTTATCTTCATGCTCTCTTGATTTTGG	420	QY
361	GGGTGAGACGGATTAAATATCATCCATGAGAGCTTTATCTTCATGCTCTCTTGATTTTGG	420	Db
421	TTTCAGATCATCTTTTCAGTGTTTCAAGAAATTTTCTCAGTTTGGTCCATGTAATTTTGG	480	QY
421	TTTCAGATCATCTTTTCAGTGTTTCAAGAAATTTTCTCAGTTTGGTCCATGTAATTTTGG	480	Db
481	AAGTGAGGTTCTTAAATTCATTATGCTTCCTTTCTTTCTAGACTAGCAACTGCATCA	540	QY
481	AAGTGAGGTTCTTAAATTCATTATGCTTCCTTTCTTTCTAGACTAGCAACTGCATCA	540	Db
541	CTTTTCACCTTGGGTTCACAAAATGTACTCAAGAAAAAAATATCACTTTTGGGTTTCA	600	QY
541	CTTTTCACCTTGGGTTCACAAAATGTACTCAAGAAAAAAATATCACTTTTGGGTTTCA	600	Db
601	AAATTCCTCTTCAGGATGTACTTTTCACTTTGAACTGTCATGTATAGGAAACAAGAAATGGCT	660	QY
601	AAATTCCTCTTCAGGATGTACTTTTCACTTTGAACTGTCATGTATAGGAAACAAGAAATGGCT	660	Db
661	CAGTTTTTAAAGGAAACAATGTACAGATTCATTTTCAGAACTCTTTCTGGTTGGTTGAGTTT	720	QY
661	CAGTTTTTAAAGGAAACAATGTACAGATTCATTTTCAGAACTCTTTCTGGTTGGTTGAGTTT	720	Db
721	CAGACTTTTTGTACCAAGCTGATGGATCAAAATCTTGTTTCCAAAGTCTGATACAGAA	780	QY
721	CAGACTTTTTGTACCAAGCTGATGGATCAAAATCTTGTTTCCAAAGTCTGATACAGAA	780	Db
781	ACTGGCAACTCTTAATTTGATAATAAAGNATAAATACAGATACAGATATCTCATTTTC	840	QY
781	ACTGGCAACTCTCTAAATTTGATAATAAAGNATAAATACAGATACAGATATCTCATTTTC	840	Db
841	TTGGTTTGGCAGATCACAAAAAGGAAACAAGGCTAAGCCTCCTACTTGTTCGGGAGTTA	900	QY
841	TTGGTTTGGCAGATCACAAAAAGGAAACAAGGCTAAGCCTCCTACTTGTTCGGGAGTTA	900	Db
901	GGTCAGGGACACATATGAATGAAGAAGAAATCTTTAAATTTGGGGTTCACCAAGATTTGTCTC	960	QY
901	GGTCAGGGACACCATATGAATGAAGAAGAAATCTTTAAATTTGGGGTTCACCAAGATTTGTCTC	960	Db
961	TCTCGAGGTTGGGGGTCCTTAGGTTGGTAGCAATACCAATATATACCTTAACAA	1020	QY
961	TCTCGAGGTTGGGGGTCCTTAGGTTGGTAGCAATACCAATATATACCTTAACAA	1020	Db
1021	ACCCAATCCATGCTACATACATACATAGCATCCATCACTTTGTAGCTGGACCTTCATCA	1080	QY
1021	ACCCAATCCATGCTACATACATACATAGCATCCATCACTTTGTAGCTGGACCTTCATCA	1080	Db
1081	AGAGCAACCATGGAGGAGCTCACATACAGCGGGCGAGCCATCGCCATTCTTCCCATAG	1140	QY
1081	AGAGCAACCATGGAGGAGCTCACATACAGCGGGCGAGCCATCGCCATTCTTCCCATAG	1140	Db
1141	CAGGGCCTTCAAGTACATCGGCTCCTCTGGTTGTCTCTCATGATCCTGTGTCAGA	1200	QY
1141	CAGGGCCTTCAAGTACATCGGCTCCTCTGGTTGTCTCTCATGATCCTGTGTCAGA	1200	Db
1201	GETTGGACCTGAGGAAGCAGAAAGCCGAGATCATGCGCAGTTCATCGGTCAACCGTGG	1260	QY
1201	GETTGGACCTGAGGAAGCAGAAAGCCGAGATCATGCGCAGTTCATCGGTCAACCGTGG	1260	Db
1261	AGCAGCTGAGGAAGCTACACCGGATGCACGACTGGCTTGTGGGTACTCTGTCACGGCACA	1320	QY

1261	AGCAGCTGAGGAACTACCAACCGATGCAAGATGGCTTGTGGGTACCTGTCCGGCACCA	1320
1321	GGACAGTGCAGCTGCAGATGCCGTTCACTTCTACACCTACATCGCTGACCCGGTGAATG	1380
1321	GGACAGTGCAGCTGCAGATGCCGTTCACTTCTACACCTACATCGCTGACCCGGTGAATG	1380
1381	TCGAGCATGTCTCWAAGACTTAACCTACCAATTCACCCAAAGGTAAATGACCTGAACCTAC	1440
1381	TCGAGCATGTCTCWAAGACTTAACCTACCAATTCACCCAAAGGTAAATGACCTGAACCTAC	1440
1441	TGATGTTCACTTTCGGAATCAGAGCTGAAGAGCTGAATCGAATGTGCCTGAACACCGTG	1500
1441	TGATGTTCACTTTCGGAATCAGAGCTGAAGAGCTGAATCGAATGTGCCTGAACACCGTG	1500
1501	TAGGGAATCGTGTACAGATCCTACATGTGACGTGCTCCTCGGTGACGGCATCTTCAACGCC	1560
1501	TAGGGAATCGTGTACAGATCCTACATGTGACGTGCTCCTCGGTGACGGCATCTTCAACGCC	1560
1561	GACGGCAGCTGTGGAGGAAGCAGAGGAAGAAGCGGAGAGTTTCGAGTTCGCTCCCAAGAAC	1620
1561	GACGGCAGCTGTGGAGGAAGCAGAGGAAGAAGCGGAGAGTTTCGAGTTCGCTCCCAAGAAC	1620
1621	CTGAGGATTTCAAGCGCATTTGTGTTTCAGAGAGTACTCCCTGAAGCTGTCCGGGTACTTG	1680
1621	CTGAGGATTTCAAGCGCATTTGTGTTTCAGAGAGTACTCCCTGAAGCTGTCCGGGTACTTG	1680
1681	AGCCAGCATCTCAAGCAGGCAAAAGTTGTGGACATGCAGGTGAGATCACTGTCTCCCTTGC	1740
1681	AGCCAGCATCTCAAGCAGGCAAAAGTTGTGGACATGCAGGTGAGATCACTGTCTCCCTTGC	1740
1741	CATTGCCAAATGAGCATTTCAACCTGAGACACGAGAGCTACCTTCCCGATTCAGGAAC	1800
1741	CATTGCCAAATGAGCATTTCAACCTGAGACACGAGAGCTACCTTCCCGATTCAGGAAC	1800
1801	TTACATCAGAGATGACGCTGGACCTCACTCTCAAGTTGGGTTCCGGGTCCGAGATCGGCAC	1860
1801	TTACATCAGAGATGACGCTGGACCTCACTCTCAAGTTGGGTTCCGGGTCCGAGATCGGCAC	1860
1861	GCTGTCGCCGATCTCCCCGAGAACAGCTTCGGCGAGCGTTTCGATGCCGCCAACATCAT	1920
1861	GCTGTCGCCGATCTCCCCGAGAACAGCTTCGGCGAGCGTTTCGATGCCGCCAACATCAT	1920
1921	CGTCAACGCTGGGTTTATCGAACCGCTGTGTGGCGCATCAAGAGGTTCTTCCACGTCGGGTC	1980
1921	CGTCAACGCTGGGTTTATCGAACCGCTGTGTGGCGCATCAAGAGGTTCTTCCACGTCGGGTC	1980
1981	AGAGCCCTCTACGCGCAGAGCATCAAGCTCGTGGAGAGTTCACTTACAGCGTGAATCCG	2040
1981	AGAGCCCTCTACGCGCAGAGCATCAAGCTCGTGGAGAGTTCACTTACAGCGTGAATCCG	2040
2041	CCGAGGAGAGCCGAGATCGTTCGAGGCCCGGCGCAGCGGCAACAGAGAGGTTACGTGC	2100
2041	CCGAGGAGAGCCGAGATCGTTCGAGGCCCGGCGCAGCGGCAACAGAGAGGTTACGTGC	2100
2101	ACATGACGTGTTTCAGTCTTCAGTTCATCGTCTTTCGGCGGATGGACCTGATCTCGATTG	2160
2101	ACATGACGTGTTTCAGTCTTCAGTTCATCGTCTTTCGGCGGATGGACCTGATCTCGATTG	2160
2161	ATTATATATCCGTGTGACTTGTGAGGACAAATTAATTAATGGCAGATGAAGCAGACATCC	2220
2161	ATTATATATCCGTGTGACTTGTGAGGACAAATTAATTAATGGCAGATGAAGCAGACATCC	2220
2221	TGTCACGGTTTCATCGAGCTAGCGAGGCGCGCGACGACGCGCGGCTTCGGGGACGACA	2280
2221	TGTCACGGTTTCATCGAGCTAGCGAGGCGCGCGACGACGCGCGGCTTCGGGGACGACA	2280
2281	AGAGCCTCCGGGACGTGTGTCTAACTTTCGTGATCGCGCGGCGGACACGACGCGCAG	2340
2281	AGAGCCTCCGGGACGTGTGTCTAACTTTCGTGATCGCGCGGCGGACACGACGCGCAG	2340
2341	CGCTGTCGTGGTTCAACGACATGCCATGTCCTCCACCGGACGTCGGCGAGAGCTGCGCC	2400

Db 2341 CGCTGTGCTGTTACGCAATGTCCTCCACCCGAGCTGGCCGAGAAAGCTGCGCC 2400  
Qy 2401 GCAGAGTGTGCGGCTTCGAGCGGAGCGCGCGCGAGGAGGCGTTCGCGCTGTCGCT 2460  
Db 2401 GCAGAGTGTGCGGCTTCGAGCGGAGCGCGCGCGAGGAGGCGTTCGCGCTGTCGCT 2460  
Qy 2461 GCAGCGGCTGAGCGCGAGCGAGGAGGCTTCGCGCGCGCGGCTTCGCGGCGC 2520  
Db 2461 GCAGCGGCTGAGCGCGAGCGAGGAGGCTTCGCGCGCGCGGCTTCGCGGCGC 2520  
Qy 2521 TCCTCACTACGAGCTCGGCAAGCTGCTACCTCCAGCGCTGCTACCGAGAGCG 2580  
Db 2521 TCCTCACTACGAGCTCGGCAAGCTGCTACCTCCAGCGCTGCTACCGAGAGCG 2580  
Qy 2581 TCCGCTGTATCCCGCGCGCTCCCTCAGGTGAGCGCGCGCGAGCAACGAGCTTCGAG 2640  
Db 2581 TCCGCTGTATCCCGCGCGCTCCCTCAGGTGAGCGCGCGCGAGCAACGAGCTTCGAG 2640  
Qy 2641 GCACAGATGAGTGGAGCTGAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 2700  
Db 2641 GCACAGATGAGTGGAGCTGAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 2700  
Qy 2701 CCCAAGGGATCTCGGAGGAGAGCTGCTCGCGAGCGGAGCAAGGTGAGCGCGCGCGG 2760  
Db 2701 CCCAAGGGATCTCGGAGGAGAGCTGCTCGCGAGCGGAGCAAGGTGAGCGCGCGCGG 2760  
Qy 2761 ATGTTGAGTACGTGCTCCCTACTCGATGCGGCGGATGAGTACAACTGCGGCGCGCG 2820  
Db 2761 ATGTTGAGTACGTGCTCCCTACTCGATGCGGCGGATGAGTACAACTGCGGCGCGCG 2820  
Qy 2821 GCGAGTCTCCGCGCGAGCGGTGATCAACAGAGTGGCGGTTCCGCAACGCGTCCGCG 2880  
Db 2821 GCGAGTCTCCGCGCGAGCGGTGATCAACAGAGTGGCGGTTCCGCAACGCGTCCGCG 2880  
Qy 2881 TTCAAGTTTCAAGGCTTCAGCGCGGCGCGAGGATCTGCTGGGCAAGGATCTGCGGCTAC 2940  
Db 2881 TTCAAGTTTCAAGGCTTCAGCGCGGCGCGAGGATCTGCTGGGCAAGGATCTGCGGCTAC 2940  
Qy 2941 CTGAGATGAAGATGCGCTGCGCATCTCTGCGCTTCTACAGCTTCGCGTCTGAG 3000  
Db 2941 CTGAGATGAAGATGCGCTGCGCATCTCTGCGCTTCTACAGCTTCGCGTCTGAG 3000  
Qy 3001 GCGCACCGGCTGAGTACCGGATGATGACCATCTCTCCATGGCGCAGGCTCAAGTTC 3060  
Db 3001 GCGCACCGGCTGAGTACCGGATGATGACCATCTCTCCATGGCGCAGGCTCAAGTTC 3060  
Qy 3061 GCGGCTCTTAGGCGCTGATGATGATGCGGATTTGGGATATCATCCGCTTAATCCCTTA 3120  
Db 3061 GCGGCTCTTAGGCGCTGATGATGATGCGGATTTGGGATATCATCCGCTTAATCCCTTA 3120  
Qy 3121 AAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3180  
Db 3121 AAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3180  
Qy 3181 CTTTAAATCTCGTGGGCTTCGGAACACCAATCACTAGTCTTTGATCTACTCT 3240  
Db 3181 CTTTAAATCTCGTGGGCTTCGGAACACCAATCACTAGTCTTTGATCTACTCTACTCT 3240  
Qy 3241 CTCAGTGAAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3300  
Db 3241 CTCAGTGAAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3300  
Qy 3301 CGGATCTTCCGCGGACCTTTTGGAGACCATTAATCTGACAGGCGGTGTGAAAAAAGGCTTC 3360  
Db 3301 CGGATCTTCCGCGGACCTTTTGGAGACCATTAATCTGACAGGCGGTGTGAAAAAAGGCTTC 3360  
Qy 3361 TTCTGCGGCAAGTCTTGGGTTGAGAGTCTTGGGCTTTGAGAGAGAGAGAGTTCG 3420  
Db 3361 TTCTGCGGCAAGTCTTGGGTTGAGAGTCTTGGGCTTTGAGAGAGAGAGAGTTCG 3420  
Qy 3421 AAGGATCTGACCTGAAACCGAAATGGCTTCGGAATATGCTCGCATCGGCGCGGCGC 3480  
Db 3421 AAGGATCTGACCTGAAACCGAAATGGCTTCGGAATATGCTCGCATCGGCGCGGCGC 3480

Qy 3481 GTCACTCGGATGACGACAGCCCAAGCAGTGAGAGCGAAGCATCTTTGGAGTTGG 3540  
Db 3481 GTCACTCGGATGACGACAGCCCAAGCAGTGAGAGCGAAGCATCTTTGGAGTTGG 3540  
Qy 3541 AGACACTCTCGGACCCCTCGCGCTCGCGAGCTCATCTTCGCTCTCTGTGTGTCG 3600  
Db 3541 AGACACTCTCGGACCCCTCGCGCTCGCGAGCTCATCTTCGCTCTCTGTGTGTCG 3600  
Qy 3601 TGGCGGACCGCGCGCGCGCTCGTGTTCGACCAATCCGCGCGCGGACCGGTTGCT 3660  
Db 3601 TGGCGGACCGCGCGCGCGCTCGTGTTCGACCAATCCGCGCGCGGACCGGTTGCT 3660  
Qy 3661 GTACAAACCTCTCATCCGCGCGCGCTCGTGTTCGACCAATCCGCGCGCGGATATACAT 3720  
Db 3661 GTACAAACCTCTCATCCGCGCGCGCTCGTGTTCGACCAATCCGCGCGCGGATATACAT 3720  
Qy 3721 CTATAATCATGTGTATTTGATCTTTATTTTCAACCGGCTTTAAACAAACCATATTTATG 3780  
Db 3721 CTATAATCATGTGTATTTGATCTTTATTTTCAACCGGCTTTAAACAAACCATATTTATG 3780  
Qy 3781 GTAAACACGTTTCAAAATTTGACAAATTTAAACAGGACCAACCGTAGCTTAAACATAAG 3840  
Db 3781 GTAAACACGTTTCAAAATTTGACAAATTTAAACAGGACCAACCGTAGCTTAAACATAAG 3840  
Qy 3841 AGAATGAGAGACCAACCCAAAGGTTAGAGATGAAATTAAGCTGAGTAAACGACGAATTC 3897  
Db 3841 AGAATGAGAGACCAACCCAAAGGTTAGAGATGAAATTAAGCTGAGTAAACGACGAATTC 3897

## RESULT 2

US-10-412-000-7  
; Sequence 7, Application US/10412000  
; Publication No. US2003018269A1  
; GENERAL INFORMATION:  
; APPLICANT: ALBERTSEN, MARC C.  
; APPLICANT: FOX, TIM  
; APPLICANT: HUFFMAN, GARY  
; APPLICANT: TRIMNELL, MARY  
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES MEDIATING MALE FERTILITY AND  
; TITLE OF INVENTION: METHOD OF USING SAME  
; FILE REFERENCE: 1148  
; CURRENT APPLICATION NUMBER: US/10/412,000  
; CURRENT FILING DATE: 2003-04-11  
; PRIOR APPLICATION NUMBER: US/09/670,153  
; PRIOR FILING DATE: 2001-06-11  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 7  
; LENGTH: 3897  
; TYPE: DNA  
; ORGANISM: Zea mays  
US-10-412-000-7

Query Match 100.0%; Score 3897; DB 14; Length 3897;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 3897; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAATTCGAAGCGAGCGCTTTAGCAGAGAGTGTTCGTGATGCGAGTTCGCGCGGAATGAGT 60  
Db 1 GAATTCGAAGCGAGCGCTTTAGCAGAGAGTGTTCGTGATGCGAGTTCGCGCGGAATGAGT 60  
Qy 61 GCGTCTCAGAGCAACGCTGAGGGTTCCAGGGATGCGCAATGGTATGCGCAATCGGCTAG 120  
Db 61 GCGTCTCAGAGCAACGCTGAGGGTTCCAGGGATGCGCAATGGTATGCGCAATCGGCTAG 120  
Qy 121 AGGTGAGGACCAAGTGTGTGAGGATTGGAGGCGCAACCTATGGCAAGTTGGTGAAGAGGC 180  
Db 121 AGGTGAGGACCAAGTGTGTGAGGATTGGAGGCGCAACCTATGGCAAGTTGGTGAAGAGGC 180  
Qy 181 ACCCAATGAGAGATCTATTTCAGACTTACATGGATGCGGCCCAACAAATTCACCTTTAGA 240  
Db 181 ACCCAATGAGAGATCTATTTCAGACTTACATGGATGCGGCCCAACAAATTCACCTTTAGA 240













Db 379 CTGTCNCGGATCTCCCGGAGACAGACTTCNCCCAAGCGTTTCGATGCGGCTAACATCATC 438  
QY 1922 GTACGCTGGGTTTCATCGACCCGCTGTGGCGCATCAAGAGTTCTTTCCAC 1972  
Db 439 GTACACCTGGGTTTCATCCACCNCCTGTGGCGCATCCAGAAGTTCTTTCCCC 489

RESULT 8  
US-10-412-000-3  
; Sequence 3, Application US/10412000  
; Publication No. US20030182689A1  
; GENERAL INFORMATION:  
; APPLICANT: ALBERTSEN, MARC C.  
; APPLICANT: FOX, TIM  
; APPLICANT: HUFFMAN, GARY  
; APPLICANT: TRIMNELL, MARY  
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES MEDIATING MALE FERTILITY AND  
; FILE REFERENCE: 1148  
; CURRENT APPLICATION NUMBER: US/10/412,000  
; CURRENT FILING DATE: 2003-04-11  
; PRIOR APPLICATION NUMBER: US/09/670,153  
; PRIOR FILING DATE: 2001-06-11  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 494  
; TYPE: DNA  
; ORGANISM: Sorghum sp.  
; FEATURE:  
; NAME/KEY: modified base  
; LOCATION: (1)..(494)  
; OTHER INFORMATION: "n" bases may be a, t, c, g, other or unknown  
US-10-412-000-3

Query Match 6.9%; Score 268.4; DB 14; Length 494;  
Best Local Similarity 77.9%; Pred. No. 3.8e-65;  
Matches 367; Conservative 0; Mismatches 28; Indels 76; Gaps 1;

QY 1502 AGGGAATCGGTGACAGATCTTACATGACGCTGCTCTCGGTGACGGCATCTTCAAGCCG 1561  
Db 95 AGGGGACGTGTACAGATCTTACATGATGTCTCTCGGTGACGGCATCTTCAAGCTG 154  
QY 1562 ACGCGAGCTGTGGAGAACACAGAGAACGCGAGTTTCGAGTTTCGCTTCCCAAGAAC 1621  
Db 155 ACGCGAGCTGTGGAGAACACAGAGAACGCGAGTTTCGAGTTTCGCTTCCCAAGAAC 214  
QY 1622 TGAGGGAATTCAGCGCATTTGTTCAGAGAGTACTCCCTGAAGCTGTGGGTACTGA 1681  
Db 215 TGAGGGAATTCAGTGCCATTTGTTCAGAGAGTACTCCCTGAAGCTGTGGGTACTGA 274  
QY 1682 GCGAGCATCCAAAGCGAGGAAAGTTGTGACATGACAGTTCGAGTCTCTCCCTTGC 1741  
Db 275 GTCAGGCATCCAAAGCGAGGAAAGTTGTGACATG 309  
QY 1742 ATTGCCACATGAGCATTTCAACTGACACAGAGCTACCTTGCAGATTCAGGAATT 1801  
Db 310 318

QY 1802 TACATGAGATGACGCTGGACTCATCTGCAAGGTTTGGGTTGGGTCGAGATCGGCAAG 1861  
Db 319 TACATGAGATGACACTGGACTCATCTGCAANGTTGGGTTGGGTCGAGATTCGCGCAG 378  
QY 1862 CTGTCGCGGATCTCCCGAGAACAGTTTCGCGAGCGGTTTCGATGCGGCAACATCATC 1921  
Db 379 CTGTCNCGGATCTCCCGAGAACAGCTTCNCCCAAGCGTTTCGATGCGGCTAACATCATC 438  
QY 1922 GTACGCTGGGTTTCATCGACCCGCTGTGGCGCATCAAGAGTTCTTTCCAC 1972  
Db 439 GTACACCTGGGTTTCATCCACCNCCTGTGGCGCATCCAGAAGTTCTTTCCCC 489

RESULT 9

US-10-021-657-6  
; Sequence 6, Application US/10021657  
; Publication No. US20020083483A1  
; GENERAL INFORMATION:  
; APPLICANT: ALBERTSEN, MARC C.  
; APPLICANT: FOX, TIM  
; APPLICANT: HUFFMAN, GARY  
; APPLICANT: TRIMNELL, MARY  
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES MEDIATING MALE FERTILITY AND  
; FILE REFERENCE: 1148  
; CURRENT APPLICATION NUMBER: US/10/021,657  
; CURRENT FILING DATE: 2001-12-14  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 267  
; TYPE: DNA  
; ORGANISM: Zea mays  
US-10-021-657-6

Query Match 6.7%; Score 263; DB 13; Length 267;  
Best Local Similarity 100.0%; Pred. No. 8.3e-64;  
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 830 ATCTATTTTCTTGTGGCAGATCACAAAAGGACACAAAGGCTTAAGCTTCTACTTG 889  
Db 5 ATCTATTTTCTTGTGGCAGATCACAAAAGGACACAAAGGCTTAAGCTTCTACTTG 64  
QY 890 TTCGGAGTTAGTTCAGGACACCATATGAATCAAGAAATCTTAATTTGGGTCACACC 949  
Db 65 TTCGGAGTTAGTTCAGGACACCATATGAATCAAGAAATCTTAATTTGGGTCACACC 124  
QY 950 AAGATTGTCTCTCGAGTTGGGGGTCCTTAAGGTTGGTAGTACCAATACCCAATATA 1009  
Db 125 AAGATTGTCTCTCGAGTTGGGGGTCCTTAAGGTTGGTAGTACCAATACCCAATATA 184  
QY 1010 TCACCTAAACACCCCAATCCATGCTACATACATACATACATACATACATACATAC 1069  
Db 185 TCACCTAAACACCCCAATCCATGCTACATACATACATACATACATACATACATACAT 244  
QY 1070 ACCCTTCATCAAGACACCATCG 1092  
Db 245 ACCCTTCATCAAGACACCATCG 267

RESULT 10  
US-10-412-000-6  
; Sequence 6, Application US/10412000  
; Publication No. US20030182689A1  
; GENERAL INFORMATION:  
; APPLICANT: ALBERTSEN, MARC C.  
; APPLICANT: FOX, TIM  
; APPLICANT: HUFFMAN, GARY  
; APPLICANT: TRIMNELL, MARY  
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES MEDIATING MALE FERTILITY AND  
; FILE REFERENCE: 1148  
; CURRENT APPLICATION NUMBER: US/10/412,000  
; CURRENT FILING DATE: 2003-04-11  
; PRIOR APPLICATION NUMBER: US/09/670,153  
; PRIOR FILING DATE: 2001-06-11  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 267  
; TYPE: DNA  
; ORGANISM: Zea mays  
US-10-412-000-6

Query Match 6.7%; Score 263; DB 14; Length 267;  
Best Local Similarity 100.0%; Pred. No. 8.3e-64;  
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 830 ATCTCATTTTCTTGTGGCAGATCAACAAGGAAACACAAAGGCTTACGCTTCTACTTG 889  
|||  
Db 5 ATCTCATTTTCTTGTGGCAGATCAACAAGGAAACACAAAGGCTTACGCTTCTACTTG 64  
|||  
QY 890 TTCCGGAGTTAGTCTAGGAGACACCATATCAATGAAGAAATCTTAATTTGGGGTCAACC 949  
|||  
Db 65 TTCCGGAGTTAGTCTAGGAGACACCATATGAATGAAGAAATCTTAATTTGGGGTCAACC 124  
|||  
QY 950 AAGATTGTTCTCTCGAGTTGGGGGTCCTTAAGTTGGTAGTAGCAATACCCAAATATA 1009  
|||  
Db 125 AAGATTGTTCTCTCGAGTTGGGGGTCCTTAAGTTGGTAGTAGCAATACCCAAATATA 184  
|||  
QY 1010 TCACCTTAACAACCCCAATCCATGCTACATACATACATACATACATACATACATACAT 1069  
|||  
Db 185 TCACCTTAACAACCCCAATCCATGCTACATACATACATACATACATACATACATACAT 244  
|||  
QY 1070 ACCCTTCATCAAGAGACCATGG 1092  
|||  
Db 245 ACCCTTCATCAAGAGACCATGG 267  
|||

## RESULT 11

US-10-425-114-23429  
; Sequence 23429, Application US/10425114  
; Publication No. US2004003488A1  
; GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128

; SEQ ID NO 23429

; LENGTH: 1807

; TYPE: DNA

; ORGANISM: Zea mays

; FEATURE:

; OTHER INFORMATION: Clone ID: LIB3596-027-C6\_FLI

US-10-425-114-23429

Query Match 4.2%; Score 164.2; DB 12; Length 1807;  
Best Local Similarity 67.6%; Pred. No. 4.2e-35;  
Matches 248; Conservative 0; Mismatches 113; Indels 6; Gaps 1;

QY 2693 CCGAGACCCCAAGGGATCTGGAGGACGAGTGCTGCCGACGGGACGAGAGGTGAGGG 2752  
|||  
Db 1213 CTCTGGATAACAAGCAGTCTTCGGAGACGAGTCTGCTGCTGACGGTTCAGCGTCAGCA 1272  
|||  
QY 2753 CCGGGGATGTGAGTACGTCCTTACGATGGGCGGATGGATCAACTGGGCC 2812  
|||  
Db 1273 AGGAGACATGTTGTTACGTCCTTACGTCCTTACGTCCTTACGTCCTTACGTCCT 1332  
|||  
QY 2813 CCGAGCGGAGCTTCCGGCGGAGCGGTGGATCAACAGGATGGCGGTTCGGCAAGC 2872  
|||  
Db 1333 AGGAGCGGAGGTTCGGCGGAGCGGTGGATCAACAGGATGGCGGTTCGGCAAGC 1392  
|||  
QY 2873 CGTCCCGTTCAAGTTCAAGGTTCCAGCGGGGCGGAGATCTGCTTGGCTTCTACAGTCCGGC 2932  
|||  
Db 1393 AGAGCCCGTTCAAAATTCAGAGGTTCCAGGCTGTCAGGATATGCTTGGGAGGAGT 1452  
|||  
QY 2933 CGCGGTACCTCAGATGAGATGGCGCTGCCATCTCTTGGCTTCTACAGTCCGGC 2992  
|||  
Db 1453 TCGCGTACCGGAGATGAAGATCTTCGCGCGGCTGCTCTCCGATTTCTTGTGTTCCGGC 1512  
|||  
QY 2993 TCGTGGAGGGCAC-----CCGGTSCAGTACCGATGATGACATCTCTCTCCATGGGC 3046  
|||

Db 1513 TCCGCGAGCGGACAAAGGAGCGGTGAATACCAGGACCATGATCAGCTCCACATCGAGC 1572  
QY 3047 ACGGCT 3053  
|||  
Db 1573 AGGGTCT 1579  
|||

## RESULT 12

US-10-425-114-35451

; Sequence 35451, Application US/10425114  
; Publication No. US2004003488A1

; GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128

; SEQ ID NO 35451

; LENGTH: 825

; TYPE: DNA

; ORGANISM: Zea mays

; FEATURE:

; OTHER INFORMATION: Clone ID: UC-ZMROB73033D09\_FLI

US-10-425-114-35451

Query Match 4.0%; Score 154.2; DB 12; Length 825;  
Best Local Similarity 65.1%; Pred. No. 1.6e-32;  
Matches 244; Conservative 0; Mismatches 128; Indels 3; Gaps 1;

QY 2698 GACCCCAAGGGATCTCGAGGACGAGTCTGCCGAGCGGACGAGAGGTGAGGCGCGC 2757  
|||  
Db 210 GACCAACAAGGAGTCTCGAGGACGAGTCTGCCGAGCGGACGCTGCTGAAGAAGGC 269  
|||

QY 2758 GGGATGTTGAGTACGTACGTGCTCTACTCGATGGGCGGATGGATCAACTGGGGCCCCGAC 2817  
|||  
Db 270 ACCAAGTGTACTACGCCATGTACTCCATGGGCGGATGGAGATCTGGGGCCAGC 329  
|||

QY 2818 GCGGAGCTTCCGGCGGAGCGGTGGATCAACAGAGATGGCGCTTCCGCAACGCGTCG 2877  
|||  
Db 330 TGCCCGAGTACAAGCGGAGCGGTGGCT---CCGGAGCGGACGCTTCTGCGGAGTCC 386  
|||

QY 2878 CGGTTCAAGTTCAAGCGGTTCAGGCGGCGGCGGAGATCTGCTGGGCAAGGACTCGCG 2937  
|||  
Db 387 GCCTCAAGTTCAAGCGCTTCAACGCGGCGGCGGCGCTGTGCTCGGCAAGGACTTCGCC 446  
|||

QY 2938 TACCTGCAGATGAAGATGGCGTGGCCATCTCTTTCGCTTCTACAGCTTCGCGTCTG 2997  
|||  
Db 447 TACTACAGATGAAGTTCAACGCGCTTCCATCTCCGCGCTACCGGTCCGCGTCTGTC 506  
|||

QY 2998 GAGGGGCGCCGCGTGCATACCGCATGATGACCATCTCTCATGGCGCAGGCTCAAG 3057  
|||  
Db 507 GAGGGCCACCCGTCGCGCCCAAGATGGCGTCCACATGTACATGAAGCAGGCGTCAAG 566  
|||

QY 3058 GTCGCGTCTTAGG 3072  
|||  
Db 567 GTGACGCTCACAAG 581  
|||

## RESULT 13

US-10-260-238-1236

; Sequence 1236, Application US/10260238  
; Publication No. US20040016025A1

; GENERAL INFORMATION:

; APPLICANT: Budworth, Paul R.  
; APPLICANT: Moughamer, Todd G.  
; APPLICANT: Briggs, Steven P.

APPLICANT: Cooper, Bret  
APPLICANT: Glazebrook, Jane  
APPLICANT: Goff, Stephen A.  
APPLICANT: Katagiri, Fumiaki  
APPLICANT: Kreps, Joel  
APPLICANT: Provart, Nicholas  
APPLICANT: Ricke, Darrell  
APPLICANT: Zhu, Tong  
TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION  
FILE REFERENCE: 60111-NP  
CURRENT APPLICATION NUMBER: US/10/260,238  
CURRENT FILING DATE: 2002-09-26  
PRIOR APPLICATION NUMBER: US 60/325,448  
PRIOR FILING DATE: 2001-09-26  
PRIOR APPLICATION NUMBER: US 60/325,277  
PRIOR FILING DATE: 2001-09-26  
PRIOR APPLICATION NUMBER: US 60/370,620  
PRIOR FILING DATE: 2002-04-04  
NUMBER OF SEQ ID NOS: 6077  
SEQ ID NO 1236  
LENGTH: 1617  
TYPE: DNA  
ORGANISM: Oryza sativa  
FEATURE:  
NAME/KEY: N region  
LOCATION: (1326)..(1326)  
OTHER INFORMATION: n = any nucleotide  
US-10-260-238-1236

Query Match 3.9%; Score 151.6; DB 15; Length 1617;  
Best Local Similarity 62.7%; Pred. No. 1.5e-31;  
Matches 235; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

QY 2697 GCACCCCAAGGGATCTCGAGGACGACGTCTCCGGACGGGACGAAGTGAGGCGGG 2756  
DB 1218 GGAGCACAAGCGCGGTTCGCCGACGACGTCTCCCGACGGGACGAGGTGATGGCGG 1277

QY 2757 CGGATGTGACGTACGTGCCCTACTCGATGGGCGGATGAGTACACTGGGCCCCGA 2816  
DB 1278 CGACAGGTGTGTGTCTTACTACTCCATGGGAGGATGAAGCGGGTNTGGGCAAGGA 1337

QY 2817 CGCGCGAGCTTCGGCGGAGCGGTGGATCAAGAGGATGGCGGTTCGCAACGCGTC 2876  
DB 1338 CTGACGGAGGTTCAGGCGGAGCGGTGGATCAACGAGGACGGGACGTGCGTACGTGCC 1397

QY 2877 GCGGTTCAAGTTACGCGGTTCCAGGCGGCGGCGGAGGATCTGCTTGGGCAAGGACTCGGC 2936  
DB 1398 GTCGAACAAGTTCTGCGGTTCAACTCGGCGCGGAGGACATGCTCGGCAAGGAGATGGC 1457

QY 2937 GTACTCGAGATGAAGATGGCGCTGGCCATCTCTTGGCTTCTACAGCTTCGCGCTGCT 2996  
DB 1458 GCTGTCGAGTGAAGTCAAGCGGCGGCGGCGGATGGGTGGAACTTTCGCGTGGAGTGGT 1517

QY 2997 GGAGGGCACCGGTGAGTACCGCATGATGACCATCTCTCCATGGGCGACGCGCTCAA 3056  
DB 1518 GCCGGGCGACGTCTGGAGCGAGGCTTCGCTATCTACTCCATGAGATGGGCTCTT 1577

QY 3057 GGTCCGCGTCTCTAG 3071  
DB 1578 GGTAAAGGTCAAGAG 1592

RESULT 14  
US-10-260-238-4619  
Sequence 4619, Application US/10260238  
Publication No. US20040016025A1  
GENERAL INFORMATION:  
APPLICANT: Budworth, Paul R.  
APPLICANT: Moughamer, Todd G.  
APPLICANT: Briggs, Steven P.  
APPLICANT: Cooper, Bret  
APPLICANT: Glazebrook, Jane  
APPLICANT: Goff, Stephen A.

RESULT 15  
US-10-425-114-28465  
Sequence 28465, Application US/10425114  
Publication No. US20040034888A1  
GENERAL INFORMATION:  
APPLICANT: Liu, Jindong  
APPLICANT: Zhou, Yihua  
APPLICANT: Kovalic, David K.  
APPLICANT: Screen, Steven E.  
APPLICANT: Tabaska, Jack E.  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53313)B  
CURRENT APPLICATION NUMBER: US/10/425,114  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 73128  
SEQ ID NO 28465

APPLICANT: Katagiri, Fumiaki  
APPLICANT: Kreps, Joel  
APPLICANT: Provart, Nicholas  
APPLICANT: Ricke, Darrell  
APPLICANT: Zhu, Tong  
TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION  
FILE REFERENCE: 60111-NP  
CURRENT APPLICATION NUMBER: US/10/260,238  
CURRENT FILING DATE: 2002-09-26  
PRIOR APPLICATION NUMBER: US 60/325,448  
PRIOR FILING DATE: 2001-09-26  
PRIOR APPLICATION NUMBER: US 60/325,277  
PRIOR FILING DATE: 2001-09-26  
PRIOR APPLICATION NUMBER: US 60/370,620  
PRIOR FILING DATE: 2002-04-04  
NUMBER OF SEQ ID NOS: 6077  
SEQ ID NO 4619  
LENGTH: 854  
TYPE: DNA  
ORGANISM: Triticum aestivum  
US-10-260-238-4619

Query Match 3.8%; Score 146.4; DB 15; Length 854;  
Best Local Similarity 61.6%; Pred. No. 2.8e-30;  
Matches 234; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 2693 CGCAGGACCCCAAGGGGATCTCGAGGACGACGTCTGCCGACGGGACGAAGGTGAGGG 2752  
DB 56 CGATGAGCACAAGGGCGTGGTGGCGGAGCGCTGCCGAGCGGSCACGAGGTGCGGC 115

QY 2753 CCGCGGGATGTTGACGTAGTCCCTACTCGATGGGCGGATGGAGTCAACTGGGGCC 2812  
DB 116 CGGGGACAAAGATCATGTGTCTGTACGCGATGGGAGGATGGAGCGGTGTGGGGCA 175

QY 2813 CCGAGCGGGCGAGCTTCGCGCGGAGCGGTGGATCAACGAGGATGGCGCTTCGCAACG 2872  
DB 176 AGACTGCGGGAGTTCCGCGCGGAGCGGTGGATCGGGAGGACGGCAAGCGCGGTACG 235

QY 2873 GGTCCGCGTTCAAGTTACGCGGTTCCAGGCGGGCGGAGGATCTGCTTGGGCAAGGACT 2932  
DB 236 TGCGCTGTAACAAGTTCTGTCTTCAACTCCGCGCGGAGATGCTCTCGCAAGGACA 295

QY 2933 CGCGGTACCTGACATGAAGATGGCGTGGCCATCTCTTTCGCGTCTTACAGCTTCGGC 2992  
DB 296 TGGCGTTCTGTCAGTCAAGCGGTGGCGCGCGCGCTGGTGGAGAACTTCGAGGTGGAG 355

QY 2993 TGTGGAGGGGACCGCGGTGACGTACCGCATGATGACCATCTCTCCATGGCGCACGCGC 3052  
DB 356 CCGTGGCGGACACGCTGAGGACCCCAAGATCTCCATCATCTCTCCATGAAGAACGGCT 415

QY 3053 TCAAGGTTCGCGTCTCTAGG 3072  
DB 416 TCAAGCGCAGGATCAAGAG 435

```
; LENGTH: 1947
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB4757-030-F2_FLI
US-10-423-114-28465

Query Match      3.7%; Score 143.8; DB 12; Length 1947;
Best Local Similarity 52.1%; Pred. No. 2.8e-29;
Matches 402; Conservative 0; Mismatches 352; Indels 17; Gaps 3;

QY 1764 CTGAGACACGAGAGCTACCTTCCCGAATTCAGAACTTTACATGAGGATGACGCTGGACT 1823
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1824 CCATCTGCAAGGTTGGGTTCCGGGTCGAGATCGGACGCTGTCGCGGATCTCCCGGAGA 1883
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1884 ACAGC---TTCCGCGAGGGTTTCGATGCCGCCAACATCATCTACGCTGCGGTTTCATCG 1940
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1941 ACCCGCTGTGGCCATCAAGAGGTTCTTCCAGTCCGGTCAGAGCCCTCTCTAGCGCAGA 2000
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2001 GCATCAAGCTCGTGGACGAGTTCACTACAGGTCGATCCGCCGAGGAGGCCGAGATCG 2060
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2061 TCGAGGCCCGGCGCAGCGGCAAAACAGGAGAGGTACGTGCACATGACTGTTTCGATTCTT 2120
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2121 CAGTTTCATCTGCTGCGCGGATGGAAGCTGATTCGATGATGATGATGATGATGATGAT 2180
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2181 GTGAGGACAAATTAATGGCAGATGAGCAGCATCTCTGTCACGGTTTCATCGAGCTA 2240
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2241 GCGAGGCGCGGCGAGCGCGGGCTTCGGGACGACAGAGCCTCCGGGACGTGGTG 2300
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2301 CTCAACTTCGTGATCGCGGCGGACACGACGGGACGACGCTGTGTTGTTTACGAC 2360
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2361 ATGCCCATGTCCACCGGAGCTGGCGAGAGCTGCGCGGAGCTGCGGCTTCGAG 2420
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2421 GCGAGGCGCGCGCGAGGAGGGGCTGCGGCTCGCTGCGGCGGCGCTGACGCGGAC 2480
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2481 GACAAAGGCTTCGCGCGCGGCGGAGTTCCGGGCTTCCTCACCTAC 2531
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2531 AGCAAGATGCACTACTGTCACGCGGCGCTGACGAGAGCCCTCCGCGCTCTAC 1281
```

Search completed: March 7, 2004, 19:35:33  
Job time : 910 secs



QY 573 GATGACGCTGGACTCATCGAAGGTTGGGTTCCGGGTCGAGATCGGCACGCTGTCCGC 632  
Db 543 CTTACCTTCGACAAATCTCGCGCTGGCTTCGGCAAGGACCCCGAGACGCTCCGCCA 602  
QY 633 AGATCTCCCGAGAACAGCTTCGCGCAGGCTTCGATGCGCGCAACATCATCATCGCT 692  
Db 603 GGGCTTCGCGAGAACAGTTCCCTCCGCTTCGACCGCGCCACGAGGCCACGCTCAA 662  
QY 693 GGGTTTCATGACCCG-----CTGTGGGCGATCAAGAGGTTCTTCCAGTGGGTCAA 746  
Db 663 CCGCTTCATCTCCCGGAGTTCTGTGGGCTGCAAAAGTGGCTGGGCTCGCATGA 722  
QY 747 GGGCTTCATGACGACAGATCAAGCTTCGTGACGAGTTCACTACAGAGTGTATCGCGC 806  
Db 723 GACCAAGCTGACAGCAGATGCGCCACATCGACAGTACTCTCGCGCGCTCATCAAGAA 782  
QY 807 GAGGAAGCGGAGATCGTCAAGGTCGCGGCCACGCGCAACAGGAGAGTGAAGCAGCA 866  
Db 783 GGGCAAGCTCGAGCTCGCGCGCGCAACGCAATCGACACGCGCGGCGACGACGACA 842  
QY 867 CATCTGTACAGTTTCATGAGTGGCGAGCGCGCGAGCAGCGCGGCGGCTTCGGGA 926  
Db 843 CTTGCTCTCCCGGTTATCGGAAGGTTCTTACTCGGACG----- 883  
QY 927 CGATAAGAGCTCCGGGAGTGTGTCTCAATCTCGTATCGCGCGCGGACACGACGCG 986  
Db 884 -----AGTCGCTCACAGCTGGCGCTCACTTCATCTCGCGCGCGGACACCTCTC 938  
QY 987 GAGGAGCTGTGTGTTCAAGCAATGCGCATGTCCACCGCGAGCTGGCCGAGAGCT 1046  
Db 939 CTTGGCGCTCTCTGTGTTCTTGTGCTGTCTTCCACCCACCTGCGGTGGAGCGAAGAT 998  
QY 1047 GCGCGCGAGCTGTGCGCTTCGAGCGGAGCGCGCGCGGAGGCGGTCACTGCT 1106  
Db 999 CTTGCGGAGCTCTGTCT----- 1019  
QY 1107 GCTCTCGCGCGGCTGTAGCGCGACGACAAAGGCTTTCGCGCGCGGTGGCGAGTTCG 1166  
Db 1020 TCTCGCGGCTCACGGGGCGCCATGACCGGCACT-----GTGGCTGGC 1064  
QY 1167 GGGCTTCCTACTAGCAGCTCGGCAAGTGTCTACTCTCACTCGCTGCTGCTCAACCGA 1226  
Db 1065 GGAGCGCTTCACTTTCGAGGAGCTCACCGCGCTGTCTACTTCAAGCGCGCGCTGTGCGA 1124  
QY 1227 GACGCTCCGCTGTACCGCGCTCTCTCAGGACCCCAAGGGATCTCTGAGGAGAGCT 1286  
Db 1125 GACCTTCGCTCTACCTCTCGTCCCGAGGACTCAAAGCAGTGTCTGCGCGAGACTA 1184  
QY 1287 GTCGCGGACGGAAGAGTGAAGCGCGCGGAGTGTGAGTACGTGCCCTACTCTCGAT 1346  
Db 1185 CTTCCCGGACGCGACCTTCTGTCGCGCGCGGTCTGCTGCTCACTTCCATATCTCGGC 1244  
QY 1347 GGGCGGATGAGTACAACTGGGCGCCCGACGCGGGGAGCTTCGCGCGGAGCGGTGAT 1406  
Db 1245 GGGCGGATGAGAGGGGTTGGGGGAGGACTGCTTCGAGTTTCGCGCGGAGCGATGGCT 1304  
QY 1407 CAACGAGGATGGCG-----GTTCCGCAACGCGTCGCGCTTCAAGTTTCAAGGCTTCGAGC 1463  
Db 1305 GTTCGCGGACGCGACCAAGTTTGAGCAGCAGCACTGTCAAGATTGTGTGGCGTTCAAAGC 1364  
QY 1464 GGGCGGAGATCTGCTGGGCAAGGATCGGCGTCACTGAGATGAAGATGGCGCTGGC 1523  
Db 1365 CGGGCGAGGGTGTGCTGGGCAAGGACCTGACCTTACCTGAGATGAAGAACATCGCGCG 1424  
QY 1524 CATCTCTTCGCTTCTAGCTTCGCGCTCGGCTGTGAGGGGACCGCGTGTGAGTACCGGAT 1583  
Db 1425 GAGCGTGTCTCTCCGCGACCGCTGACCGTGGCGCGCGGCGGCGGCGTGGAGAGAGAT 1484  
QY 1584 GATGACCATCTCTCCATCGCGCAGGGCT 1613  
Db 1485 GTTCGCTACGCTCTTCAAGGGCGGGCT 1514

## RESULT 2

US-09-158-767-10  
; Sequence 10, Application US/09158767A  
; Patent No. 6180363  
; GENERAL INFORMATION:  
; APPLICANT: Batard, Yannick  
; APPLICANT: Durst, Francis  
; APPLICANT: Schalk, Michel  
; APPLICANT: Werck-Reichhart, Daniele  
; TITLE OF INVENTION: RECODING OF DNA SEQUENCES PERMITTING  
; TITLE OF INVENTION: EXPRESSION IN YEAST AND OBTAINED TRANSFORMED YEAST  
; FILE REFERENCE: A32000  
; CURRENT APPLICATION NUMBER: US/09/158,767A  
; CURRENT FILING DATE: 1998-09-23  
; EARLIER APPLICATION NUMBER: FR 97-12094  
; EARLIER FILING DATE: 1997-09-24  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 10  
; LENGTH: 2181  
; TYPE: DNA  
; ORGANISM: *Triticum aestivum*  
US-09-158-767-10

Query Match 14.2%; Score 271.6; DB 3; Length 2181;  
Best Local Similarity 54.5%; Pred. No. 4.6e-47;  
Matches 736; Conservative 0; Mismatches 524; Indels 90; Gaps 6;

QY 276 CACCTACATCGCTGACCCGCTGAATGTGAGCATGTCTCAAGACTAACTTACCACATTA 335  
Db 354 CACCGTCACTCGACCCCGCGCACTTGGAGCAGCTCTTGAAGCGCGCTTCGACACTA 413  
QY 336 CCCAAGGATTCGTGTACAGATCTCAATGAGCGTGTCTCTGTGAGCGGATTTCAA 395  
Db 414 CCCCAGGCGCTCTTGGCACGGCTTTCGGGACCTGTCTGGGAGCGGATCTTCAA 473  
QY 396 CGCGACCGGAGCTGTGGAGGAGCAGAGAGCGGAGTTCGAGTTGCGCTC--- 452  
Db 474 TTCCAGCGGCGACACTGGCTCGCGCAGCGACGCGCGCTCGAGTTTCAACCCG 533  
QY 453 CAAGAACCTGAGGAGTTTTCAGCGCATTTCTCAGAGAGTACTTCCCTGAAGCTCTCGG 512  
Db 534 CACGCTCCGAGCGGCAATGTCGCTGGGTCTCGCGCTCCATCCAGCGCGCTCTGCG 593  
QY 513 TATCTGACCGAGGATCCAGGCGAGGCAAGTGTGTGAGCATGAGGAACTTTACATGAG 572  
Db 594 CATCTGCGCGACGCGGCAAGGCGCAGGTGTATCTCCAGGACCTCTCTCTCGG 653  
QY 573 GATGACGCTGACTCCATCTCAGAGTTGGGTTCGGGCTCGAGATCGGACGCTGCGCC 632  
Db 654 CTTACCTTCGACACATCTCGGGCTGGCTTCGGCAGGACCGGAGAGCGCTCGCCCA 713  
QY 633 AGATCTCCCGAGAACAGCTTCGCGCAGGCGTTCGATSCCGCCCAACATCATCACGCT 692  
Db 714 GGGCTCGCGAGAACAGTTGCTCGCGCTTCGACCGCGCCACCGAGGCCACGCTCAA 773  
QY 693 GGGTTTCATCAACCG-----CTGTGGCGCATCAAGGTTCTTCCAGTGGGTGAGA 746  
Db 774 CCGTTTCATCTTCCCGGAGTTTCTCTGTGGCGCTGCAAAAGTGGCTGGGCTCGCATGGA 833  
QY 747 GGGCTCTCTAGCGCAGAGCATCAAGCTCGTGAGAGTTCACCTACAGCGTGTACCGCG 806  
Db 834 GACACGCTGACGAGAGCATGGCCACAGTCCAGTACCTCGCGCGCTCATCAAGAA 893  
QY 807 GAGGAAGCGGAGATCGTCAAGGTCGCGGCGAGCGGCAACAGGAGAGATGAAGACGA 866  
Db 894 GCGCAAGCTCGAGCTCGCGCGCGCAACGCAATGCGACGCGCGGCGGACGACGACA 953  
QY 867 CATCTGTACGCTTCTGAGCTGGGAGGCGCGGACGACGAGCGGCGGCTTCGGGGA 926  
Db 954 CTTCTCTCCCGTTCTATGCGGAAGGTTCTTACTCGGACG----- 994  
QY 927 CGATAAGAGCTCCGGGAGCGTGTGCTCAATCTGTGATCGCGCGGCGGACGACGCGC 986

Db 995 -----AGTGGCTCCAGACGCTGGCGCTCACTTCATCTCGCCGGCGGACACCTCCCTC 1049  
QY 987 GACGAGCTCTGTGGTTCACGACATGGCCATGTCCACCCGGAGTGTCGCCCGAGAGCT 1046  
Db 1050 CGTGGGCTCTCTGTGGTTCCTTCTGGCTCGTGTGTCCACCCACCTGCGGTGAGCGCAAGAT 1109  
QY 1047 GCGCCGCGAGCTGTGGCGTTCGAGGCGGAGCGCGCGCGGAGGCGCTCACGCTCGT 1106  
Db 1110 CGTGGCGAGCTGTGCTC-----CGT 1130  
QY 1107 GCTCTCGCGCGGCGCTGACCGCGACGACAAAGGGTTCGCGCCGGTGGCGAGTTCCG 1166  
Db 1131 TCTCGCGCGCTACGCGGCGCCCATGACCCGGCATT-----GTGGTGGC 1175  
QY 1167 GCGCCTCTCTACCTACGACAGCTCGGACAGCTGTGTCTACCTCAGCGCTGCGTCAACGA 1226  
Db 1176 GGAGCCCTTTCACCTTCGAGGAGTCGACCGCTGTGTCTTACCTCAAGCGCGCGCTGTGGA 1235  
QY 1227 GAGCCTCGGCTCTACCCCGCGTTCCTCAGGACCCCAAGGGATCTTGAGGACGAGCT 1286  
Db 1236 GACCTTCGGCTCTACCTTCCTCGTCCGCGCGGTGCTGACGACTCAAGCAGCTGTCGCGACGACTA 1295  
QY 1287 GTCGCGGAGCGGACGAAGGTGAGGCGCGCGCGGATGGTACGCTACGTGCTGCTTACTCGAT 1346  
Db 1296 CTTCCCGGAGCGGACCTTGTGTCGCGCGCGGTGCTGCTCACTTCCATATACTCGG 1355  
QY 1347 GGGGCGGATGGAGTCAACTGGGCGCCCGACGCGCGGAGCTTCCGCGCGGAGCGGTGGAT 1406  
Db 1356 GGGGCGGATGAAGGGGTGTGGGCGGAGGACTGCTCGAGTTCCGCGCGGAGCGATGGCT 1415  
QY 1407 CAACGAGGATGGGCG---GTTCCGCAACGCGTCCGCGTTCAAGTTACGCGGTTCCAGGC 1463  
Db 1416 GTGCGCGGACGACCAAGTTTCAGCAGCAGCAGCTGTTACAGTTGTCGCGTTCAACGC 1475  
QY 1464 GGGGCGGAGATCTGCTGGGCAAGACTCGGCTACCTGCGATGAAGATGGCGTGGC 1523  
Db 1476 CGGCGCGGAGGTGTGCTGGGCAAGGACTAGCTTACCTGCGATGAAGACATCGCGG 1535  
QY 1524 CATCTCTTCGCTTACAGTTCGCGTGTGAGGAGGCGACCCGCGTCACTGCGAT 1583  
Db 1536 GAGCGGTGCTGCTCGGCAACGCTGACCGTGGCGCGCGGCGCACCGGTTGAGCAGAAGAT 1595  
QY 1584 GATGACCATCTCTCCATGCGCGACGCGCT 1613  
Db 1596 GTGCTCACGCTTTCATGAGGCGGGCT 1625

## RESULT 3

US-09-249-585A-4/c

; Sequence 4, Application US/09249585A

; Patent No. 6417002

; GENERAL INFORMATION:

; APPLICANT: Horlick, Robert

; TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISOMES

; FILE REFERENCE: 0867/0905

; CURRENT APPLICATION NUMBER: US/09/249,585A

; CURRENT FILING DATE: 1999-02-11

; NUMBER OF SEQ ID NOS: 16

; SOFTWARE: Patent version 3.0

; SEQ ID NO 4

; LENGTH: 1926

; TYPE: DNA

; ORGANISM: Epstein Barr Virus

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (1)..(1926)

; OTHER INFORMATION: template strand of EBNA-1 DNA

US-09-249-585A-4

Query Match

Best Local Similarity 3.8%; Score 73; DB 4; Length 1926;

Matches 345; Conservative 0; Mismatches 410; Indels 6; Gaps 2;

QY 737 TGGGTTCAGAGGCCCTCTTAGCGCAGAGCATCAAGCTCGTGGACGAGTTCACCTACAGCG 796  
Db 1014 TGGGGCCGAGGTGACGAGGAGCTTGGGGCCGAGGAGTGGAGGACGCGGAGGACG 955  
QY 797 TATCCGCGCGAGGAGGCGGAGATCGTGAAGTCCGGGCCAGCGCAACAGAGAAGA 856  
Db 954 AGGACGGGAGGAGGACGAGGACGGGAGGACGGGAGGACGGGAGGACGGGG 895  
QY 857 TGAAGCAGACATCTCTGTTCACGCTTCATCAGCTGGCGAGGCCGCGACGACGGCGGG 916  
Db 894 AGGACGAGGACGGGAGGACGGGAGGACGAGGACGGGAGGACGGGAGGACGAG 835  
QY 917 GCTTCGGGACGATAGACCTTCCGGGACGTGGTGTCTCACTTCGTGATCGCGGCGGG 976  
Db 834 ACGGAGGAGACGGGAGGAGGAGGACGAGGACGGGAGGAC-----GGGAGGACGAGGAC 780  
QY 977 ACAACGAGCGGACGAGCTGTCTGTGTTTCAACACATGGCCATGTCCACCCCGGACGTCG 1036  
Db 779 GGAGGACGGGAGGACGAGGACGGGAGGACGGGAGGACGAGGAC-GGGAGGAGGACGAG 721  
QY 1037 CCGAGAGCTGCGCCCGGAGCTGTGCGGTTTCAGGCGGAGCGCGCGCGGAGGAGGCG 1096  
Db 720 ACGGGAGGACGAGGACGGGAGGACGAGGACGGGAGGACGGGAGGACGGGAGGAG 661  
QY 1097 TCACGCTGTCTCTGCGCGCGGCTGACCGCGACGACGAGGCTTCCGCGCGCGCGGTGG 1156  
Db 660 ACGAGGACGGGAGGACGAGGACGGGAGGACGGGAGGACGGGAGGACGAGGACGGGG 601  
QY 1157 CGCAGTTCCGCGGCTCTCTACCTACGACGAGCTCGGCAAGCTGTCTACCTCCAGCT 1216  
Db 600 AGGAGGACGAGGACGGGAGGACGGGAGGACGGGAGGACGAGGACGGGAGGAGCAG 541  
QY 1217 GCGTCAACGAGAGCTGCGCTTACCCCGCTTACCCCGCTTCCCTCAGGACCCCAAGGGAG 1276  
Db 540 GGGAGGAGGACGAGGACGGGAGGAGGACGAGGACGGGAGGACGGGAGGACGGGAGG 481  
QY 1277 AGGACGAGCTGTCTGCGGACGCGGACGAGGTCGAGGTCGCGCGCGGATGCTGAGCTAG 1336  
Db 480 AGGACGAGGACGGGAGGACGGGAGGAGGACGAGGACGGGAGGAGGACGAGGACGGGG 421  
QY 1337 CCTACTCGATGGGCGGATGAGTACAACTGGGCGCCCGGACGCGGAGCTTCCGCGCGG 1396  
Db 420 AGGACGGGAGGACGGGAGGAGGACGAGGACGGGAGGACGGGAGGACGGGAGGAGCAG 361  
QY 1397 AGCGTGGATCAACGAGGATGCGGCTTCCGCAACGCGTTCGCGCTTCAAGTTTACGGGCT 1456  
Db 360 GGGAGGACGGGAGGACGGGAGGAGGACGAGGACGGGAGGAGGACGAGGACGGGAGG 301  
QY 1457 TCCAGCGGGCGCGGAGGATCTGCTTGGGCAAGGACTCGGCG 1497  
Db 300 ACGAGGACGGGAGGCGGAGGACGAGGACGAGGACGAGGACGAGGACGAGGAGG 260

## RESULT 4

US-09-130-114-2/c

; Sequence 2, Application US/09130114

; Patent No. 5976807

; GENERAL INFORMATION:

; APPLICANT: Horlick, Robert A.

; APPLICANT: Damaj, Bassem B.

; APPLICANT: Robbins, Alan K.

; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes

; TITLE OF INVENTION: From Multiple Transfected Episomes

; FILE REFERENCE: 0867/1D903US1

; CURRENT APPLICATION NUMBER: US/09/130,114

; CURRENT FILING DATE: 1998-08-06

; NUMBER OF SEQ ID NOS: 36

; SOFTWARE: FastSeq For Windows Version 3.0

; SEQ ID NO 2

; LENGTH: 1931

; TYPE: DNA

; ORGANISM: EBNA

US-09-130-114-2

Query Match 3.8%; Score 73; DB 2; Length 1931;  
Best Local Similarity 45.3%; Pred. No. 3.3e-06;  
Matches 345; Conservative 0; Mismatches 410; Indels 6; Gaps 2;

QY 737 TCGGGTCAGAGCCCTCCTAGCGCAGAGCATCAAGCTGCTGACGAGTTTCACTACAGG 796  
DB 1014 TGGGCGGAGGTGACGAGGAGCTGGGCGCGAGGTGAGGACGAGGAGCGGAGGACG 955  
QY 797 TGATCCGCGGAGGAGGCGGAGAGCTGTCGAGGTCGCGGCCAGCGGCCAAACAGGAGAGA 856  
DB 954 AGGACGGGAGGAGGACGAGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 895  
QY 857 TGAAGCAGACATCTGTACAGTTTATCGAGTGGGAGGAGCGGCGGAGGAGGAGGAGG 916  
DB 894 AGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 835  
QY 917 GTTCCGGGAGGATAGAGCCTCCGGGAGCTGTGCTCAACTTCTGTGATCGCGGCGGG 976  
DB 834 ACGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 780  
QY 977 ACACAGCGCGACGAGCTGTGCTGTTTACGACATGSCCATGTCCACCCCGGAGCTGG 1036  
DB 779 GGAGGACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 721  
QY 1037 CCGAGAAGCTGGCGCGGAGCTGTGCGGCTTCGAGGCGGAGGCGCGCGGAGGAGGCGG 1096  
DB 720 ACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 661  
QY 1097 TCACGCTGTGCTCTCGCGCGGCGCTGACCGCGAGCAAGGAGCTTCGCGCGCGCGTGG 1156  
DB 660 ACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 601  
QY 1157 GCGATTTCGCGGCGCTCTCTACCTACGAGCCTCGGCAAGCTGTCTACCTCCAGCCT 1216  
DB 600 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 541  
QY 1217 GCGTACCGAGAGCTCGCGCTGTACCGCGCTCCCTCAGGACCCGAGGGATCTTGG 1276  
DB 540 GGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 481  
QY 1277 AGGAGCAGCTGTGTCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1336  
DB 480 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 421  
QY 1337 CTTACTCGATGGGCGGATGAGATCAACTGGGGGCGGAGCGCGGCGGAGCTTCGCGCGG 1396  
DB 420 AGGACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 361  
QY 1397 AGCGTGTGATCAACAGGATGCGGCTTCGCGCAACGCTGCGCTCAAGTTTCAAGGCT 1456  
DB 360 GGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 301  
QY 1457 TCCAGCGGGGCGGAGGATCTCTCTGGCAAGGACTCGCGG 1497  
DB 300 ACGAGGACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 260

## RESULT 5

US-09-252-991A-289/c  
; Sequence 289, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 289

; LENGTH: 1365

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-289

Query Match 3.7%; Score 70.6; DB 4; Length 1365;  
Best Local Similarity 46.5%; Pred. No. 9.6e-06;  
Matches 229; Conservative 0; Mismatches 264; Indels 0; Gaps 0;

QY 753 CCTAGCGCAGAGCATCAAGCTCTGAGCGAGCTTCACTACAGCGTTCATCCGCGCGAGGAA 812  
DB 985 CCACGCCGAACGCTACAGGCTCGGTGGAATTCACCGCATCTGGCGTCCGCTGCTGGA 926  
QY 813 GGCAGAGATCTCGAGTTCGCGGCCAGCGGCAACAGAGAGATGAAGCAGCAGCATCCT 872  
DB 925 GGGCAGAGCCGCTCGACTACGCGCGCAAGCACATCAGGTGAAGGCGCCAACTGCTCTA 866  
QY 873 GTCAGGTTTATCGAGCTGGCGGAGGCGGCGAGCAGACGCGCGCGGCTTCGGGGACGATAA 932  
DB 865 TCCGCGCTTGACGAGCGCGCTCCGCGCTGTACTTCGCGGCTCTCTCGAGGCGCGCCA 806  
QY 933 GAGCTCCGGGACGCTGTGCTCAACTTCGTGATCGCGGGCGGACAGCAGCGGCGAGCAG 992  
DB 805 GGACCTCCCGCGGAGGAGTTCGAGCTGTACCTGAGCTGGGGCGAGCGCGCGGCGGCT 746  
QY 993 GCTGTCTGTGTTACGACATGGGCTATGTCACCGCGACGTCGCGCGAGAGAGCTGCGCG 1052  
DB 745 TGCCGAGAAGATGCGCCAGGTACCGAGAGGCGCGCGCGCGCGCGCGCGCGCTT 686  
QY 1053 CGAGCTGTCGCTTCGAGGCGGAGCGCGCGGCGGAGAGGCGTCAAGCTCTGCTGCTG 1112  
DB 685 CGGATCCGCTTCGACGCTGATCGTCCGCGAGACCGAGGAGGAGGAGGAGGAGGAGG 626  
QY 1113 CGCGCGGCTGACGCGCGAGCAAGGCTTCGCGCGCGCGCTGCGAGTTTCGCGGGCT 1172  
DB 525 TCGACTGATCGCGCACTTCGAGCAGACACCATCGCGCGCGCGCGCGCGCGCTT 566  
QY 1173 CTTCACTTACAGAGCTTCGCAAGCTGCTTACCTCAAGCTTCGCTACCGAGAGCGCT 1232  
DB 565 CTTCACTTTCGCGCGAGCAACGATGCGCGCGCTGACGCGGCTTCGCGCGAGCAACT 506  
QY 1233 CCGCTCTACCCC 1245  
DB 505 GGAAGTCAGCCCC 493

## RESULT 6

US-09-252-991A-248  
; Sequence 248, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 248  
; LENGTH: 1941  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-248

Query Match 3.7%; Score 70.6; DB 4; Length 1941;  
Best Local Similarity 46.5%; Pred. No. 1e-05;

Matches	229;	Conservative	0;	Mismatches	264;	Indels	0;	Gaps	0;
QY	753	CCTAGCGCAGACGATCAAGCTCGTGGACGAGCTTCACCTACAGCGTGATCCGCGGAGAA	812						
Db	1164	CCACGCCGAACGCTACGAGGCTCGGTGCAATTCACCGGATTCGGCGGTGCGGTCTGGG	1223						
QY	813	GGCCGAGATCGTCSAGGTCGGGCCACGCGCAACAGGAGAAGATGAAGCACGACATCCCT	872						
Db	1224	GGCGGACACCGTCGACTACGCCGCAAGCACATCCAGGTGAAGGGCGCCAACTGCTCTA	1283						
QY	873	GTACAGGTTTCATCGAGCTGGCGGAGCGCGGACGACGGCGCGGCTTCGGGACGATAA	932						
Db	1284	TCCGCGGTGACAGACGCGGTCCGCGCGCTGTACTTCGGCGGCTCTTCGGAGCGCGCCA	1343						
QY	933	GAGCTCCGGACGTGTGCTCAACTTCGTGATCGCGGGGGGACACGACGGCGACAC	992						
Db	1344	GGACCTCGCGCGACGACGTTCGACTGTACCTTGGGGGAGCGCGCGCGCGGT	1403						
QY	993	GCTGTCTGTGTTACGACACATGGCCATGTCTCCACCCGAGCGTGCCGAGAAAGTGTGSCCG	1052						
Db	1404	TGCCGAGAAGATGCCCCAGGTACCGGAGAGGCGCCCGCCAGGGCCGCGAGTGGCTT	1463						
QY	1053	CGACTGTGGCGTTTCAGGCGGAGCGCGCGCGAGAGGGCGTTCACGCTCGTGTCTG	1112						
Db	1464	CGGGATCCGCGCTGCACCTGATCGTCCGCGAGACGAGAGAACCTTGGCAGCGCCCGA	1523						
QY	1113	CGGCGGCGCTGACGCCGACACACAGGCGTTGCGCGCCCGCGTGGCGAGTTCGCGGCGCT	1172						
Db	1524	TGACTGATCGCGCACTTCACGACGACACCAATCGCCCGCGCCAGGCTCTCCCTGGCGG	1583						
QY	1173	CCTCACCTACGACAGCCTTCGCAAGCTTGGTCTACCTCCAGCGCTGCGTCAACCGAGCGT	1232						
Db	1584	CTTCGACTCGTGGCCACGAAACGATGGCGCGGTGCACGGCGGTTTCGCGCACACCT	1643						
QY	1233	CGCGCTGTACCC	1245						
Db	1644	GGAAGTCAGCCCC	1656						

RESULT 7  
US-09-252-991A-238  
; Sequence 238, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

	Query Match	3.7%;	Score 70.6;	DB 4;	Length 2076;
	Best Local Similarity	46.5%;	Pred. No. 1.1e-05;		
	Matches 229;	Conservative 0;	Mismatches 254;	Indels 0;	Gaps 0;
QY	753	CTTAGCGCAGCATCAAGCTCGTGGACGAGTTCACCTACAGCGTGATCCGCGGAGAA	812		
Db	490	CCACGGCGAACGCTACAGGCGCTCGGTGCGAATCACCGGCATCTGGCGTGCSTGTGGA	549		
QY	813	GGCGGAGATCGTTCGAGTCCGGGCGCGCAACAGGAGAGATGAAGCACGACATCCT	872		
Db	550	GGCGGAGACGCTGCATACGCCGGCAGACATCCAGTGAAGGGCGCAAACTGCTCTA	609		

QY 873 GTCA CGGTTCATCGAGCTGGGCGAGGCGGCGGACGACGCGGCGGGCTTCGGGACGATAA 932

Db 610 TCGCCCGTTGCAGCAGCGCGGTCCGCGCGTGTACTTCGGCGGCTCCTCGAGGCGGCCCA 669

QY 933 GAGCCTCCGGAGCGTGGTGTCTAACTTCGTGATCGCGCGGCGGGACACGACGCGACGAC 992

Db 670 GGA CCTCGCGCGGAGCAGGTTCGAGCTGTACCTGTGACTGTGGGCGAGCGCGCGGCGGT 729

QY 993 GCTGTCTGTGTTCA CGCACATGGCCATGTCCCA CCGGAGCTGCGCGAGAGCTGCGCGG 1052

Db 730 TGCCGGAAGATCGCC CAGGTACGCGAGAGGCGCGCCGAGGCGCCAGGTGCGCTT 789

QY 1053 CGAGCTGTGCGGTTTCGAGCGGAGCGCGCGCGAGGAGGCGTCA CGCTCGTCTCTG 1112

Db 790 CGGGATCCGCGCTCAGCTGATCTCGCGAGACGACGAGAACTGGCAGCGCCCGA 849

QY 1113 CGCGGCGCTGACGCGCAGACAAAGCGTTCGCGCGCGCGGTGGCGCAGTTCGCGGCGCT 1172

Db 850 TCACTGATCGGCACCTCGACGACACCATCGCGCGCGCCAGGCGCTCCCTGGCGG 909

QY 1173 CCTCACTACGACAGCGCTCGCAAGCTGGTCTACTTCCACGCTCGGTCA CGGACGCT 1232

Db 910 CTTTCGACTCGGTGGCGCAGCAACGATGCGCGCGCTGCACGGCGTTCGCGGACAACT 969

QY 1233 CGCGCTGTACCC 1245

Db 970 GGAAGTCAGCCCC 982

RESULT 8

US-09-252-991A-5309

US-09-252-991A-5309 : Sequence 5309, Application US/09252991A

Patent No. 6551795 : GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 5309

LENGTH: 1551

TYPE: DNA

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-5309

	Query Match	3.6%;	Score 69.2;	DB 4;	Length 1551;
	Best Local Similarity	45.6%;	Pred. No. 1.9e-05;		
	Matches 282;	Conservative 0;	Mismatches 333;	Indels 3;	Gaps 1;
QY	689	CGCTGGGTTTCATCGACCCGCTCTGGCGCATCAAGAGTTCTTCACGTCGGGTACAGG	748		
Db	324	CGCAGTGATCTCAACCAAGGTACCGGGCGCAACCGCAGCACCTCGCGCGGTACACCG	383		
QY	749	CCCTCTCTACGCGAGAGCATCAAGTCTGTGGACGAGTTCACCTACAGCGTGATCCGCCGGA	808		
Db	384	AGGTGGCCGGGCAGTCGGCGCGGGTGATCGTCCGCAACCGGCACGGGATCACTCTGCCAGG	443		
QY	809	GSAAAGCCGAGATCGTCTGAGGTCCGGGCCAGCGGCAAAACAGGAGAAGATGAACACAGACA	868		
Db	444	GCTGCGGTTTCATCAACACGCGCGCGCGACCTCACCAACCGCAAGCGCATCATGGACG	503		
QY	869	TCCTGTCAAGGTTTCATCGAGCTGGGCGAGCGCGGACACGCGCGCGGCTTCGGGGACG	928		
Db	504	GCACAGGCTTGAGGCGTTCCAGGTGGACGGCGGGGACATCGTCTGTCGAAGGCGCGCAAC	563		
QY	929	ATAAGAGCCTCCGGGACGTGTGTCTCAACTTCGTGATCGCGGGCGGGACACGACGCGGA	988		

Db 564 TGAACGTCGGCAACCTCGAACAGTTCGACCTGATCACCGCGCGCAAGCTCAACGCCA 623  
QY 989 CGACGCTGTGTTGTTACGACATGGCCATGTCCACCCCGGACGT---GGCCGGAAGC 1045  
Db 624 AGCTCTACGCGAAGAACCTCAACATCGTCAACCGCGCAACGACGTCCAGGCGGACAGCC 683  
QY 1046 TGGCGCGGAGCTGTGCGGTTTCGAGCGGAGCGCGCGGAGGAGGCGCTGACGCTCG 1105  
Db 684 TGAAGGCCACCGCGCGCGCGATGGACGAGAGGACCAAGCTGGCGATCGACAGCT 743  
QY 1106 TGCTTCGCGCGCGCTGACCGCGACGACAAAGCGTTTCGCGCGCGCTGGCGGAGTTTCG 1165  
Db 744 CGCGCTGGCGGCGATGTACCGCGGCGATCCGCTGTTCGCGACCGGAGGCGGTGG 803  
QY 1166 CGGGCTCTCCTACCTAGGACGCTCGGCAAGCTGTCTACCTCAGCGCTGCTGTCACCG 1225  
Db 804 GGTGCGGTGGCGCGGACATGGCGCGGAGCGCGGCGGACATCCGCAATCGACGCCAGCG 863  
QY 1226 AGACGCTCCGCTGTACCCCGCGTCCCTCAGACCCCAAGGGATCCTGGAGGAGCG 1285  
Db 864 GCAAGCTGAGCTGGCGGAGCTTCAGCAGGCGGACCTGAAGATCGCGGCCAGGCGG 923  
QY 1286 TGCTGCGGACGGGACGA 1303  
Db 924 TGGAGCTGAACGGCAAGA 941

## RESULT 9

US-09-252-991A-5348  
; Sequence 5348, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 5348  
; LENGTH: 4158  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-5348

Query Match 3.6%; Score 69.2; DB 4; Length 4158;  
Best Local Similarity 45.6%; Pred. No. 2.4e-05;  
Matches 282; Conservative 0; Mismatches 333; Indels 3; Gaps 1;

QY 689 CGCTCGGTTTCATCGACCGCTGTGGCGCATCAAGAGGTTCTTCCACGTCGGGTTCAGAGG 748  
Db 404 CGAGGTGATCTCAACAGGTCACCGCGGCAACCGGAGCACCTTGGCGGCTACACCG 463  
QY 749 CCCTCTAGCGCAGACATCAAGCTCGTGGACGAGTTTACCTACAGCTGTATCCCGCGGA 808  
Db 464 AGGTGGCGGCGAGTCGGCGGGGTGATCGTGGCCAAACCCCGCAGCGCATCACTGCCAGG 523  
QY 809 GGAAGCGCGAGATCGTCGAGTTCGGCGCGGCGGCAACAGGAGAGATGAAGCAGCACA 868  
Db 524 GCTGCGGTTTCATCAACAGCGCGCGGACCTTACACCGGCAAGCGGATCATGGAGC 583  
QY 869 TCCTGTACGTTTATCGAGTGGCGGAGCGCGGCGAGCGCGCGGCTTCGGGGAGC 928  
Db 584 GCCAGCGCTGAGCGCTTCCAGTGGACCGCGGCGACATCGTCTCGAAGCGCGCGAAC 643  
QY 929 ATAAGAGCTCCGGGAGCGTGTGCTCAACTTCGTGATCGCGGGGCGGACAGCGGCGGA 988  
Db 644 TGAACGTGGCAACCTCGAACAGTTCGACCTGATCACCGGAGCGGCCCAAGCTCAACGCCA 703

QY 989 CGACGCTGTGTTGTTACGACATGGCCATGTCCACCCCGGACGT---GGCCGGAAGC 1045  
Db 704 AGCTCTACGCGAAGAACCTCAACATCGTCAACCGCGCAACGACGTCCAGGCGGACAGCC 763  
QY 1046 TGGCGCGGAGCTGTGCGGTTTCGAGCGGAGGAGCGCGCGGAGGAGGCGCTGACGCTCG 1105  
Db 764 TGAAGGCCACCGCGCGCGCGATGGACGAGAGGACCAAGCTGGCGATCGACAGCT 823  
QY 1106 TGCTTCGCGCGCGCTGACCGCGACGACAAAGCGTTTCGCGCGCGCTGGCGGAGTTTCG 1165  
Db 824 CGCGCTGGCGGCGATGTACCGCGGCGATCCGCTGTTCGCGACCGGAGGCGGTGG 883  
QY 1166 CGGGCTCTCCTACCTAGGACGCTCGGCAAGCTGTCTACCTCAGCGCTGCTGTCACCG 1225  
Db 884 GGTGCGGTGGCGCGGACATGGCGCGGAGCGCGGCGGACATCCGCAATCGACGCCAGCG 943  
QY 1226 AGACGCTCCGCTGTACCCCGCGTCCCTCAGGACCCCAAGGGATCCTGGAGGAGCG 1285  
Db 944 GCAAGCTGAGCTGGCGGAGCTTCAGCAGGCGGACCTGAAGATCGCGGCCAGGCGG 1003  
QY 1286 TGCTGCGGACGGGACGA 1303  
Db 1004 TGGAGCTGAACGGCAAGA 1021

## RESULT 10

US-09-252-991A-5227/c  
; Sequence 5227, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 5227  
; LENGTH: 4953  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-5227

Query Match 3.6%; Score 69.2; DB 4; Length 4953;  
Best Local Similarity 45.6%; Pred. No. 2.5e-05;  
Matches 282; Conservative 0; Mismatches 333; Indels 3; Gaps 1;

QY 689 CGCTCGGTTTCATCGACCGCTGTGGCGCATCAAGAGGTTCTTCCACGTCGGGTTCAGAGG 748  
Db 4595 CGCAGGTGATCTCAACAGGTCACCGCGGCAACCGGAGCACCTTGGCGGCTACACCG 4536  
QY 749 CCCTCTAGCGCAGACATCAAGCTCGTGGACGAGTTTCACTACAGCTGTATCCCGCGGA 808  
Db 4535 AGGTGGCGGCGAGTCGGCGGGGTGATCGTGGCCAAACCCCGCAGCGCATCACTGCCAGG 4476  
QY 809 GGAAGCGCGAGATCGTCGAGTTCGGCGCGGCGCAACAGGAGAGATGAAGCAGCACA 868  
Db 4475 GCTGCGGTTTCATCAACAGCGCGCGGACCTTACCGCGGCAAGCGGATCATGGAGC 4416  
QY 869 TCCTGTACGTTTATCGAGTGGCGAGGCGCGGCGAGCGAGCGCGGCTTCGGGGAGC 928  
Db 4415 GCCAGCGCTGAGCGCTTCCAGGTGGACGCGGCGGACATCGTCTCGAAGGCGCGCGAAC 4356  
QY 929 ATAAGAGCTCCGGGAGCGTGTGCTCAACTTCGTGATCGCGGGGCGGACAGCGGCGGA 988  
Db 4355 TGAACGTGGCAACCTCGAACAGTTCGACCTGATCACCGGAGCGGCCCAAGCTCAACGCCA 4296  
QY 989 CGACGCTGTGCTGTTTCAGGCACATGCCCATGTCCACCCCGGACGT---GGCCGGAAGC 1045

Db 4295 AGCTCTACGGAAGAACCTCAACATGCTACCGCGCCGCAAGAGCTCCAGCGCCGACAGCC 4236  
QY 1046 TGGCCCGGAGCTGTGGCGGTTTGAGGCGAGAGCGCGCGGAGGAGGCGGTCAAGCTCG 1105  
Db 4235 TGCAGGCGACCGCGCGCGCGCGGATGGCAGCGAGAGCCACAGCTGGCGATCGACAGCT 4176  
QY 1106 TGCTCTCGCGCGCGCTGAGCGCGGACGACAAAGGGTTTCGCGCGCGCGGTGGCGAGTTGG 1165  
Db 4175 CGGCGCTGGGCGGAGTATAGCGCGGCGGATCGCTGGTGGGACCGAGACAGGCGGTGG 4116  
QY 1166 CGGCGCTCTCACTACGACAGAGCTCGGCAAGTGGTCTACCTCCAGCGCTGGTCAACCG 1225  
Db 4115 GGGTGGCGCTGGCGCGGACATGGCGCCAGCGCGCGGACATCGGATCGAGCGCCAGCG 4056  
QY 1226 AGAGCTCCGCTGTACCGCGCGCTCCCTCAGAGCCCGAGGAGATCTCGAGGAGAGCG 1285  
Db 4055 GCAAGCTGAGCCCTGGCGCGGCTCCAGCGCGGCGAGCTGAAGATGGCGCGCCAGCGCG 3996  
QY 1286 TGCTGCGGAGCGGAGCGA 1303  
Db 3995 TGGAGCTGAACGCAAGA 3978

## RESULT 11

US-09-615-192A-103  
; Sequence 103, Application US/09615192A  
; Patent No. 6410718  
; GENERAL INFORMATION:  
; APPLICANT: Blockberg, Leonard N.  
; APPLICANT: Havukkala, Ilkka  
; TITLE OF INVENTION: Materials and Methods for the  
; FILE OF INVENTION: Modification of Plant Lignin Content  
; FILE REFERENCE: 11000.10034U  
; CURRENT APPLICATION NUMBER: US/09/615,192A  
; CURRENT FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 08/975,316  
; PRIOR FILING DATE: 1997-11-21  
; PRIOR APPLICATION NUMBER: US 08/713,000  
; PRIOR FILING DATE: 1996-09-11  
; PRIOR APPLICATION NUMBER: US 09/169,789  
; PRIOR FILING DATE: 1998-10-09  
; NUMBER OF SEQ ID NOS: 405  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 103  
; LENGTH: 1866  
; TYPE: DNA  
; ORGANISM: Eucalyptus grandis  
US-09-615-192A-103

Query Match 3.6%; Score 68.6; DB 4; Length 1866;  
Best Local Similarity 47.0%; Pred. No. 2.6e-05;  
Matches 352; Conservative 0; Mismatches 384; Indels 13; Gaps 4;  
QY 894 CGAGCGCGGACGAGCGCGCGGTTCGGGACGATAGAGCCTCCGGGAGCTGGTGTCT 953  
Db 705 CAAGACGAGTTCGGTGGAGCGGATCAAGATGTGATCCGACATGGTGCAGATCTGTCT 764  
QY 954 CAACCTCGTGTATCGCGCGGAGACGAGCGGAGCGCTGTGTGTTCAGCACAT 1013  
Db 765 GGCCTTCTACGCGAGCAAGAGGTGAACGAGTCCGACGATTTGAGAACTCGATCAG 824  
QY 1014 GGCCATGTCCCGCGGAGCTGGCGGAGAGTGGCGCGGAGCTGTGGCGGTTCGAGGC 1073  
Db 825 GCTACGAGAGACACATCAAGGCCATCATATGACGTGTGTTCCGCGGAGCGAGAC 884  
QY 1074 GAGCGCGCGCGAGGAGGCGGTTCAGCTGTGTCTCGCGGCGGTTCAGCGGACGA 1133  
Db 885 TGTGGCGTGGCTATCGAGTGGCCATGCGGAGCTCATGCGAGCCCGGAGACCTGNA 944  
QY 1134 CAAGGGCTTGGCGCGCGGTGGCGGAGTTTCGGGCGCTTCCTCACTAGCA-----CAG 1187  
Db 945 GAAGSTCCAGCAAGAACTCGCGGATGTCTGTGGCGCTTAGACCGGAGAGTTCGAGAGCGA 1004

QY 1188 CCTCGCAAGCTGGTCTACTCTCACGCGCTGGTCAACGAGACGCTCGGCTGTATACCCCGC 1247  
Db 1005 CTTTCGAGAGAGCTGACCTATCTCAAGTGTGCTCAAGAGAGACCTCGGCTCCACCCGCC 1064  
QY 1248 CGTCCCTCAGGACCCCAAGGGAGTCTCGAGGACGAGCTGTCTCGGAGCGGAGAGGT 1307  
Db 1065 GATCCCGCTGCTCTCCAGAGACGCGAGAGACGCGGTGATCTC---CGGTACCGCAT 1121  
QY 1308 GAGGCGCGCGGATGGTACGCTACGTGCTTACTCGATGGGCGGATGGAGTACAACTG 1367  
Db 1122 CCGCGCACGCTCCCGGTCATGATCAATGATGGGCCATCGGCG---TGACCCCGGCTC 1178  
QY 1368 GGGCGCGGAGCGGCGAGCTTCGGCGGAGGCTGGATCAAGAGGATGGCGGTTCG 1427  
Db 1179 GTGGACCGAACCTGACAAAGTTCAAACCGCTCCCGGTTCTTGGAGTCAGGATGCCGACTA 1238  
QY 1428 CAACGCGTCCGCTTCAAGTTTCAAGTTCACGGCGTTCAGGCGGCGCGAGGATCTGCTGGGCA 1487  
Db 1239 CAAGGGAGCACTTCGAGTTTATCCCTTCGGTGGCGCGGAGTCTGCTCCAGGAT 1298  
QY 1488 GGAAGTGGCTTACCTGACAGTGAAGATGGCGTGGCGCATCTTCCGCTTCTACAGCTT 1547  
Db 1299 GCAGCTCGGCTCTACGCGCTCGACATGGCGCTGGCGCCACCTCTGCTTACGCTG 1358  
QY 1548 CCGGCTGTGGAGGCGCGCGTGCAGTACCGCATGATGACCATCTCTCATGGCGCA 1607  
Db 1359 GGAAGTGGCGCGGATGAAGCGGAGATGGACATGGCGGAGCTTTC---GGGCTCA 1417  
QY 1608 CGGCTTCAAGTTCGCGTCTCTAGGCGCG 1636  
Db 1418 CCGCGCGGAGTCCACCGGCTCTGGCG 1446

## RESULT 12

US-09-169-789-103  
; Sequence 103, Application US/09169789  
; Patent No. 6653528  
; GENERAL INFORMATION:  
; APPLICANT: Blockberg, Leonard N.  
; APPLICANT: Havukkala, Ilkka  
; TITLE OF INVENTION: Materials and Methods for the  
; FILE OF INVENTION: Modification of Plant Lignin Content  
; FILE REFERENCE: 11000.1003c2  
; CURRENT APPLICATION NUMBER: US/09/169,789  
; CURRENT FILING DATE: 1998-10-09  
; EARLIER APPLICATION NUMBER: US 08/975,316  
; EARLIER FILING DATE: 1997-11-21  
; EARLIER APPLICATION NUMBER: US 08/713,000  
; EARLIER FILING DATE: 1996-09-11  
; NUMBER OF SEQ ID NOS: 185  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 103  
; LENGTH: 1866  
; TYPE: DNA  
; ORGANISM: Eucalyptus grandis  
US-09-169-789-103

Query Match 3.6%; Score 68.6; DB 4; Length 1866;  
Best Local Similarity 47.0%; Pred. No. 2.6e-05;  
Matches 352; Conservative 0; Mismatches 384; Indels 13; Gaps 4;  
QY 894 CGAGCGCGGACGAGCGCGCGGTTCGGGACGATAGAGCCTCCGGGAGCTGGTGTCT 953  
Db 705 CAAGACGAGTTCGGTGGAGCGGATCAAGATGTGATCCGACATGGTTCGACGATCTGTCT 764  
QY 954 CAACCTCGTGTATCGCGCGGAGACGAGCGGAGCGCTGTGTGTTCAGCACAT 1013  
Db 765 GGCCTTCTACGCGAGCAAGAGGTGAACGAGTTCGAGAACTCGATCAG 824  
QY 1014 GGCCATGTCCCGCGGAGCTGGCGGAGAGTGGCGCGGAGCTGTGGCGGTTCGAGGC 1073  
Db 825 GCTACGAGAGACACATCAAGGCCATCATATGACGTGTGTTCCGCGGAGCGAGAC 884

QY	1074	GGAGGGCGCGCGAGGAGCGCTCAGCTCGTCTCTCGCGCGCGCTGACGCCGACGA	1133
Db	885	TGTGGCGTTCGGTATCGAGTGGGCGATCGGCGAGCTCATCGAAGCCCCGAGGACCTGAA	944
QY	1134	CAAGGCGTTCCGCCCGCGTGGCGCAGTTCGCGGGCTCTCTACCTACGA-----CAG	1187
Db	945	GARGGTCAGCAAACTCGCGGATGTCTGGGGCTTAGACGGAGAGCTCAGGAGAGCGGA	1004
QY	1188	CCTCGGCAAGTGTCTACTCCACGCTGCGTCACCGAGAGCTCCGCTGTACCCCGC	1247
Db	1005	CTTCGAGAGCTGACCTATCTCAAGTGTGCTCTAAAGAGACCTCGCTTCACCCGCC	1064
QY	1248	CGTCCCTCAGGACCCCAAGGGATCTCTGGAGGACGAGCTGTGCCGACGGGACGAAGGT	1307
Db	1065	GATCCCGTGTCTCTCCACGAGACGGCAGAGACCGCGTGATCTC---CGGTACCGCAT	1121
QY	1308	GAGGGCGGGGATGTGAGTACGTGCGCTACTCGATGGGCGGATGGAGTACAACCTG	1367
Db	1122	CCCCGACGGTCCGGGTCTATGATCAATGATGGGCCATCGGGC---TGACCCCGGCTC	1178
QY	1368	GGGCCCGCAGCGCGCGAGCTTCCGGCCGAGCGGTGGATCAACGAGGATGGCGCTTCGG	1427
Db	1179	GTGACCGAAGCTGACAACTGACAACTGTCGCTCCCGTTCCTGGAGTACGGATGCCGACTA	1238
QY	1428	CAACGCTCGCGGTTCAAGTTTCAAGCGTTCCAGGGGGCCGAGGATCTGCTGGGCAA	1487
Db	1239	CAAGGGAGCAACTTCGAGTTTCATCCCTTTCGGGTCCGGCCGAGGTCGTGCCAGGGAT	1298
QY	1488	GGACTCGGCGTACCTGCAGATGAAGTGGCGCTGGCCATCTCTTCGGCTTCTACAGCTT	1547
Db	1299	GCAGCTCGGCTCTAACCGCTTCGACATGGCGGTGGCCCACTCTCTGACTGCTTCAGTG	1358
QY	1548	CCGGCTCTGGAGGGGACCCGGTGCAGTACCGCATGATCACCATCTCTCCATGCGGCA	1607
Db	1359	GGAATCCCGCAGGGATGAAGCCGACGAGATGACATGGGCGACGTCTTC--GGGCTCA	1417
QY	1608	CGGCTCAAGGTCGGGCTCTCTTAGGGCG	1636
Db	1418	CCGGCCGAGGTTCCACCCGGCTCGTGGCG	1446

RESULT 13

US-09-615-192A-404  
; Sequence 404, Application US/09615192A

; Patent No. 6410718

```

/ GENERAL INFORMATION:
/ APPLICANT: Bloksberg, Leonard N.
/ APPLICANT: Ravukkaia, Ilkka
/ TITLE OF INVENTION: Materials and Methods for the
/ TITLE OF INVENTION: Modification of Plant Lignin Content
/ FILE REFERENCE: 11000.1003c4U
/ CURRENT APPLICATION NUMBER: US/09/615,192A
/ CURRENT FILING DATE: 2000-07-12
/ PRIOR APPLICATION NUMBER: US 08/975,316
/ PRIOR FILING DATE: 1997-11-21
/ PRIOR APPLICATION NUMBER: US 08/713,000
/ PRIOR FILING DATE: 1996-09-11
/ PRIOR APPLICATION NUMBER: US 09/169,789
/ PRIOR FILING DATE: 1998-10-09
/ NUMBER OF SEQ ID NOS: 405
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 404
/ LENGTH: 2013
/ TYPE: DNA
/ ORGANISM: Eucalyptus grandis
US-09-615-192A-404

```

Query Match	3.6%	Score 68.6;	DB 4;	Length 2013;
Best Local Similarity	47.0%;	Pred. No. 2.7e-05;		
Matches 352;	Conservative	0;	Mismatches 384;	Indels 13; Gaps 4;

853	Db	CAAGACGAGTTCCGGTGGAGGCGATCAAGATGTCGATACCGACATGTCGACGACTGCT	912
954	Qy	CAACTTCGTGATCGCCGGGGGACAGACGGCGACGACGCTGTCGTGGTTACGCGACAT	1013
913	Db	GGCTTTCTACGCGACGAAGCAAGGTGAACAGATCCGACGATTTGCAGAACTCGATCAG	972
1014	Qy	GGCCATGTCACACCCGGACGTGGCCGAGAAAGCTCGCCGGGAGCTGCGCGGAGCTGTCGCGTTCCGAGGC	1073
973	Db	GCTAACGAGAGACAACATCAAGGCCATCATATGGAAGTATGTTTCGCGGGACGGAGAC	1032
1074	Qy	GGAGGCGCGCGCGAGGAGGCGGTCAAGCTCGTCTCTGCGGGCGGCGCTCAAGCCGACGA	1133
1033	Db	TGTGGCGTTCGGCTATCGAGTGGGCGCATGCGGGAGCTCATCGAAGCCCGGAGACCTGAA	1092
1134	Qy	CAAGGCGTTTCGCGCGCCGCTGGCGGAGTTTCGCGGGCCCTCCTACCTACGA-----CAG	1187
1093	Db	GAAAGTCCAGCAAGAACTCCGCGGATGTCGTGGCGCTAGACCGGAGAGTCCAGGAGACGA	1152
1188	Qy	CCTCGGCAAGTGTGTTCACTCCACGCGCTCGTCAACGAGACGCTCCGCTGTACCCCGC	1247
1153	Db	CTTCGAGAACTGACCTATCTCAAGTGCCTGCTCAAGAGACCTTCGCTCCACCGCC	1212
1248	Qy	CGTCCCTCAGACCCCAAGGGGATCTCTGGAGGACGAGTCTCCGGACGGGACGGAAGT	1307
1213	Db	GATCCCGCTGTCTCTCCACGAGACGGCAGAGGACGCGGTGATCTC---CGGCTACCGCAT	1269
1308	Qy	GAGGGCGGGCGGATGGTGAAGTACGTGCGCTACTCGATGGGGCGAGTGAAGTACAATG	1367
1270	Db	CCCCGACGTCGCGGGTCATGATCAATCATGGGCGCATCGGGCG---TGACCCCGGCTC	1326
1368	Qy	GGGCCCCGACGCGCGGAGCTTCGCGCGGAGCGGTGGATCAACGAGAGTGGCGGTTCCG	1427
1327	Db	GTGACCGAACCCTACAGATTCAAACGCTCCCGTTCTTGAGTCAAGCATGCCGACTA	1386
1428	Qy	CACCGCGTCGCGGTTCAAGTTCAAGCGGTTCCAGCGGGGCGGAGGATCTGCTGGGCAA	1487
1387	Db	CAAGGGGAGCAACTTCGAGTTTATCCTCTTCGGGTTCGGGCGGAGGTCGTGCCACGGAT	1446
1488	Qy	GGACTCGGCTACCTCGAGATGAAGATGGCGCTGGGCGATCCTCTTCGGCTTTCAAGCTTT	1547
1447	Db	GCAGCTCGGGCTCTACGCGCTCGACATGGCGGTGGCCACCTCTCTGCACTGTTTCACTG	1506
1548	Qy	CCGGCTGCTGGAGGGGCAACCCGCTGCAGTACCGCATGATGACCATCCTCTCCATGGCGCA	1607
1507	Db	GGAACTGCCCGACGGGATGAAGCCGAGCGAGATGGACATGGGCGAGCTTTC-GGCTCA	1565
1608	Qy	CGCGCTCAAGGTCGCGTCTCTAGGGCGG	1636
1566	Db	CCGCGCGGAGGTTCCACCCCGCTCGTGGCG	1594

RESULT 14

US-09-252-991A-15934  
; Sequence 15934, Application US/09252991A

; Patent No. 6551795

```

; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 15934
; LENGTH: 1563
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15934

```

Query Match 3.5%; Score 67.6; DB 4; Length 1563;  
Best Local Similarity 45.5%; Pred. No. 4.1e-05;  
Matches 281; Conservative 0; Mismatches 334; Indels 1; Gaps 1;

QY 689 CGTGGCGTTTCATCGACCCCGCTGTGGCGCATCAAGAGTTTCTCCAGTGGGTTCAGAGG 748  
DB 324 CGCAGGTGATCTCAACAGGTACCCGGCGCAACCCGACGACACCTTGGCGGCTACACCG 383  
QY 749 CCCTCTAGCGCAGAGCATCAGTCTGTGAGAGTTTCACTACAGGTTATCCGCGCGA 808  
DB 384 AGTGGCGGCGCATCGCGCGGTGTCTGCGCAACCCGCAAGGATCACCTGCCAGG 443  
QY 809 GGAAGCGCGAGATCGTCCAGTCCGGCGCAGCGCAACAGGAGAGTGAAGCAACGACA 868  
DB 444 GCTCGGCTTCATCAACACGCGCGCGGACCTTCAACCGGCAAGCCGATCATGGAGC 503  
QY 869 TCCTGTCAAGTTTCATCAGTGTGGCGAGCGCGCGAGCAGCGCGGCTTCCGGGAGC 928  
DB 504 GCCAGCGCTTGGAGCGCTTCCAGGTGACGCGCGGAGCATCGTCTGAAAGCGCGAAC 563  
QY 929 ATAAGAGCTCCGGGAGTGTGCTCACTTCGTGATCGCGCGGCGGACACGACGCGGA 988  
DB 564 TGAAGTGGGCAACTCGAAGTTCAGCTGATCAACCGCGCGCAAGCTCAACGCCA 623  
QY 989 CGAGCTGTCTGTGTTTCAACGACATGGCCATGTCACACCGGAGCT---GGCGGAGAGC 1045  
DB 624 AGCTTACGCGAAGAACTTCAACATGTACCGGCGCAACGAGCTCCAGCGCGACGCC 683  
QY 1046 TGGCGCGAGCTGTGGCGGTTTTCAGCGGAGCGCGCGCGAGGAGGCGTTCACGCTCG 1105  
DB 684 TGCAGGCCACGCGCGCGCGCGGATGGCAGCGAGAGGACACAGTGGCGATCGACAGCT 743  
QY 1106 TGCTCTGCGCGCGCTGACGCGGAGCGACAGAGCGTTCGCGCGCGCGGTGGCGAGTTG 1165  
DB 744 CGCGCTGGCGGGATGTACCGCGGGCGATCGCGCTGTGCGCACCGAAGCGGCGTGG 803  
QY 1166 CGGCGCTCTCTACCTACGACAGCTCGGCAAGCTGTCTTACCTCCAGCGCTGGTTCACCG 1225  
DB 804 GGGTGGCGTGGCGCGGACATGGCGCGCAGCGCGCGGAGCATCGGATCGACCGCAGCG 863  
QY 1226 AGAGCTTCGCGTGTACCGCGCTTCCAGACCGGATCGGAGGATCTTGGAGGACGAGC 1285  
DB 864 GCAAGCTGAGCTGTGGCGCGGCTCCAGCGAGCGGCGACCTGAAGATCGCGCGCGCGCG 923  
QY 1286 TGCTGCGGAGCGGAGCA 1303  
DB 924 TCGAGCTGACGCGCAAGA 941

RESULT 15  
US-09-252-991A-15763/c  
; Sequence 15763, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIORITY FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 15763  
; LENGTH: 1953  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-15763

Query Match 3.5%; Score 67.6; DB 4; Length 1953;  
Best Local Similarity 45.5%; Pred. No. 4.3e-05;

Matches 281; Conservative 0; Mismatches 334; Indels 1; Gaps 1;

QY 689 CGTGGCGTTTCATCGACCCCGCTGTGGCGCATCAAGAGTTTCTCCAGTGGGTTCAGAGG 748  
DB 1595 CGCAGGTGATCTCAACAGGTACCCGGCGCAACCCGACGACACCTTGGCGGCTACACCG 1536  
QY 749 CCCTCTAGCGCAGAGCATCAGTCTGTGAGAGTTTCACTACAGGTTATCCGCGCGA 808  
DB 1535 AGTGGCGGCGCATCGCGCGGTGTCTGCGCAACCCGCAAGGATCACCTGCCAGG 1476  
QY 809 GGAAGCGCGAGATCGTCCAGTCCGGCGCAGCGCAACAGGAGAGTGAAGCAACGACA 868  
DB 1475 GCTCGGCTTCATCAACACGCGCGCGGACCTTCAACCGGCAAGCCGATCATGGAGC 1416  
QY 869 TCCTGTCAAGTTTCATCAGAGTGGGCGAGCGCGCGGCGGCGGCTTCCGGGAGC 928  
DB 1415 GCCAGCGCTTGGAGCGCTTCCAGGTGAGCGCGCGGACATCGTCTGAAAGCGCGAAC 1356  
QY 929 ATAAGAGCTCCGGGAGTGTGCTCACTTCGTGATCGCGCGGCGGAGACACGACGCGGA 988  
DB 1355 TGAAGCTCGGCAACTCGAAGTTCGACCTGATCAACCGCGCGGCGGCAAGCTCAACGCCA 1296  
QY 989 CGAGCTGTCTGTGTTTCAACGACATGGCCATGTCACACCGGAGCT---GGCGGAGAGC 1045  
DB 1295 AGCTTACGCGAAGAACTTCAACATGTACCGCGCGCAACGAGCTCCAGCGCGACGCC 1236  
QY 1046 TGGCGCGAGCTGTGCGGTTTTCAGGCGGAGCGCGCGCGGAGGCGTTCACGCTCG 1105  
DB 1235 TGCAGGCCACGCGCGCGCGCGGATGGCAGCGAGAGCAGCTGGCGATCGACAGCT 1176  
QY 1106 TGCTCTGCGCGCGCTGACGCGGAGCGACAGAGCGTTCGCGCGCGCGGTGGCGAGTTG 1165  
DB 1175 CGCGCTGGCGGGATGTACGCGCGGCGGATCGCGCTGTGCGCACCGAAGCGGCGTGG 1116  
QY 1166 CGGCGCTCTCTACCTACGACAGCTTCGCGCAAGCTGTCTTACCTCCAGCGCTGGTTCACCG 1225  
DB 1115 GGGTGGCGTGGCGCGGACATGGCGCGCAGCGCGGCGGAGCATCGGATCGACCGCAGCG 1056  
QY 1226 AGAGCTTCGCGTGTACCGCGCTTCCAGACCGGATCGGAGGATCTTGGAGGAGCAGC 1285  
DB 1055 GCAAGCTGAGCTGTGGCGCGGCTTCCAGCGAGCGGCGACCTGAAGATCGCGCGCGCGCG 996  
QY 1286 TGCTGCGGAGCGGAGCA 1303  
DB 995 TCGAGCTGAACGCGCAAGA 978

Search completed: March 6, 2004, 02:20:54  
Job time: 158 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 7, 2004, 12:06:17 ; Search time 6189 seconds  
(without alignments)  
18803.186 Million cell updates/sec

Title: US-10-021-657-7

Perfect score: 3897

Sequence: 1 gaattccagcgagccctt.....gcgagtaaacgacgaattc 3897

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	976.8	25.1	1002	29	CG344416
2	939.2	24.1	991	29	CG047035
3	890.8	22.9	1070	29	CC889743
c 4	860	22.1	882	29	CC635363

c 5	858.6	22.0	887	29	CG067339
c 6	858	22.0	880	29	CG047034
c 7	857	22.0	879	29	CC685050
c 8	855.6	22.0	864	29	CG344408
9	844.6	21.7	915	29	CC635355
10	805.6	20.7	859	29	CG317335
c 11	802.2	20.6	887	29	CC611162
c 12	793.8	20.4	819	28	CC368220
c 13	779.2	20.0	848	29	CC723851
c 14	775.2	19.9	811	29	CG280923
c 15	774.2	19.9	997	11	AY110120
c 16	761.4	19.5	794	29	CC685043
c 17	751	19.3	773	28	CC368222
c 18	737.4	18.9	812	29	CC723843
c 19	721.6	18.5	771	29	CC610854
c 20	700.4	18.0	712	29	CG067338
21	694	17.8	733	29	CG371423
22	693	17.8	905	29	CG086520
c 23	674.4	17.3	833	29	CG317324
c 24	657.4	16.9	681	29	CG371414
c 25	654.8	16.8	748	28	BZ532472
c 26	651.6	16.7	700	29	CG086519
c 27	603.8	15.5	615	29	CG611171
c 28	512.4	13.1	517	10	AW424821
c 29	506	13.0	676	28	BZ798912
30	461.2	11.8	534	29	CG273559
31	460.4	11.8	563	10	AW519943
32	454	11.6	502	29	CG252095
33	450.2	11.6	474	28	BH872488
34	448	11.5	492	28	BH798910
c 35	389.4	10.0	419	28	BH872489
36	386	9.9	585	12	BI075273
37	385	9.9	509	12	BI099541
38	384.2	9.9	606	10	BE494080
c 39	383.6	9.8	399	29	CG368180
40	375.8	9.6	672	29	CG257531
c 41	375.4	9.6	866	29	CG252417
42	372	9.5	649	12	BG948366
43	357.6	9.2	518	12	BG948686
44	356.4	9.1	578	12	BI075441
45	344.8	8.8	720	13	BQ840694

## ALIGNMENTS

RESULT 1  
CG344416/c  
LOCUS  
DEFINITION  
CG344416 genomic survey sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Zea mays  
Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.  
REFERENCE  
AUTHORS  
1 (bases 1 to 1002)  
Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T., Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.  
TITLE  
JOURNAL  
COMMENT  
Unpublished (2002)  
Other GSSs: OG1DH18TH  
Contact: Cathy Whitelaw  
TIGR

CG344416  
OG1DH18TV ZM\_0.7\_1.5 KB Zea mays genomic clone ZMMBMA0736C12,  
1002 bp DNA linear GSS 26-AUG-2003

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TF

```

FEATURES
  source
    Class: sheared ends.
    Location/Qualifiers
      1..1002
        /organism="Zea mays"
        /mol_type="genomic DNA"
        /strain="B73"
        /db_xref="taxon:4577"
        /clone_lib="ZMMBMA0736K12"
        /note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
        methylation filtered genomic DNA library"

ORIGIN
Query Match      25.1%; Score 976.8; DB 29; Length 1002;
Best Local Similarity 99.1%; Pred. No. 1.4e-151;
Matches 993; Conservative 0; Mismatches 7; Indels 2; Gaps 1;

QY 1082 GAGCACCATGAGGAAGCTCACATCAGCGCGGACGCGCATCGCCATTCTTCCCACTAGC 1141
DB |||||
QY 1002 GAGCACCATGAGGAAGCTCACATCAGCGCGGACGCGCATCGCCATTCTTCCCACTAGC 943
DB |||||
QY 1142 AGGSCCTCACAAATACATCGCGCTCCCTCGGTGTCTCTCATGGATCCTGGTCCAGAG 1201
DB |||||
QY 942 AGGSCCTCACAAATACATCGCGCTCTTCTGGTGTCTCTCATGGATCCTGGTCCAGAG 893
DB |||||
QY 1202 GTGAGCCTGAGGAAGCAGAAAGCCCGAGATCATGCGCCAGTATCGGTGCAACGGTGA 1261
DB |||||
QY 882 GTGAGCCTGAGGAAGCAGAAAGCCCGAGATCATGCGCCAGTATCGGTGCAACGGTGA 823
DB |||||
QY 1262 GCAGCTGAGGAATACCAACGGATGACGACTGGCTTGTGGGTACTCTGTCAGGCACAG 1321
DB |||||
QY 822 GCAGCTGAGGAATACCAACGGATGACGACTGGCTTGTGGGTACTCTGTCAGGCACAG 763
DB |||||
QY 1322 GACAGTGACCGTGCACATGCGGTTCATCTCTACACCTACATCGCTGACCGGTGAATGT 1381
DB |||||
QY 762 GACAGTGACCGTGCACATGCGGTTCATCTCTACACCTACATCGCTGACCGGTGAATGT 703
DB |||||
QY 1382 CGAGCATGTCTCAAGATCAACTTACCAATACCCCAAGTAAATGACCTGAACCTCACT 1441
DB |||||
QY 702 CGAGCATGTCTCAAGATCAACTTACCAATACCCCAAGTAAATGACCTGAACCTCACT 643
DB |||||
QY 1442 GATGTTCAAGTCTCGGAATCAGAGCTGAAGCTGAATGCAATGCTGCTCAACACCGTGT 1501
DB |||||
QY 642 GATGTTCAAGTCTCGGAATCAGAGCTGAAGCTGAATGCAATGCTGCTCAACACCGTGT 593
DB |||||
QY 1502 AGGGAATCGTGTACAGATCTCTACATGACGCTGCTCTCGGTGACGGCATCTTCAACGGCG 1561
DB |||||
QY 582 AGGGAATCGTGTACAGATCTCTACATGACGCTGCTCTCGGTGACGGCATCTTCAACGGCG 523
DB |||||
QY 1562 ACGCGAGCTGTGAGGAAGCAGAGGAGACGGGAGTTCGAGTTCGCTCCAGAAC 1621
DB |||||
QY 522 ACGCGAGCTGTGAGGAAGCAGAGGAGACGGGAGTTCGAGTTCGCTCCAGAAC 463
DB |||||
QY 1622 TGAGGATTTTCAGCGCATTTGTTTCAGAGAGTACTCCCTGAAGCTGTGCGGTACTGA 1681
DB |||||
QY 462 TGAGGATTTTCAGCGCATTTGTTTCAGAGAGTACTCCCTGAAGCTGTGCGGTACTGA 403
DB |||||
QY 1682 GCAGGATCCAAAGGAGGCAAAAGTTGTGGACATGACAGGTGAG--ATCACTGCTCCCTTG 1739
DB |||||
QY 402 GCAGGATCCAAAGGAGGCAAAAGTTGTGGACATGACAGGTGAGATATCACTGCTCCCTTG 343
DB |||||
QY 1740 CCAATTGCCACATGAGCATTTCAACCTGACAGACGAGAGTACTCTTGCCTGAGTCAAGAAC 1799
DB |||||
QY 342 CCAATTGCCACATGAGCATTTCAACCTGACAGACGAGAGTACTCTTGCCTGAGTCAAGAAC 283
DB |||||
QY 1800 TTTACATGAGGATGACGCTGACTCCATCTGCAAGTTGGTTTCGGGTTCGAGATCGGCA 1859
DB |||||
QY 282 TTTACATGAGGATGACGCTGACTCCATCTGCAAGTTGGTTTCGGGTTCGAGATCGGCA 223
DB |||||
QY 1860 CGCTGTGCGCGGATCTCCCGGAGAACAGCTTCGCGGAGGGGTTTCGATGCGCGCAACATCA 1919
DB |||||
QY 222 CGCTGTGCGCGGATCTCCCGGAGAACAGCTTCGCGGAGGGGTTTCGATGCGCGCAACATCA 163
DB |||||

```

```

QY 1920 TCGTCACGCTGCGGTTCATCGACCCGCTGTGGCGCATCAAGAGTTCCTTCCACGTCGGGT 1979
DB |||||
QY 162 TCATCACGCTGCGGTTCATCGACCCGCTGTGGCGCATCAAGAGTTCCTTCCACGTCGGGT 103
DB |||||
QY 1980 CAGAGGCGCTCTTCTAGCGCAGAGCATCAAGCTCGTGACGAGTTCACCTACAGGTCATCC 2039
DB |||||
QY 102 CAGAGGCGCTCTTCTAGCGCAGAGCATCAAGCTCGTGACGAGTTCACCTACAGGTCATCC 43
DB |||||
QY 2040 GCCGAGGAGGAGCGCGAGATCGTCGAGGCGCGGCGCAGCGCA 2081
DB |||||
QY 42 GCCGAGGAGGAGCGCGAGATCGTCGAGGTCGCGGCGCAGCGCA 1

RESULT 2
LOCUS CG047035 991 bp DNA linear GSS 19-AUG-2003
DEFINITION PUFNM667D ZM 0.6_1.0_KB Zea mays genomic clone ZMMBTA0749K12,
genomic survey sequence.
ACCESSION CG047035
VERSION CG047035.1 GI:33919215
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Zukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 991)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
Bennetzen,J.
Maize Genomics Consortium
Unpublished (2003)
Other GSSs: PUFNM66TB
CONTACT: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.
FEATURES
  source
    Location/Qualifiers
      1..991
        /organism="Zea mays"
        /mol_type="genomic DNA"
        /strain="B73"
        /db_xref="taxon:4577"
        /clone="ZMMBTA0749K12"
        /clone_lib="ZM_0.6_1.0_KB"
        /note="Vector: pCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
        Cot selected genomic DNA library"

ORIGIN
Query Match      24.1%; Score 939.2; DB 29; Length 991;
Best Local Similarity 98.4%; Pred. No. 1.4e-144;
Matches 970; Conservative 0; Mismatches 13; Indels 3; Gaps 2;

QY 842 TGGTTGGCAGATCACAAAAGGAACACAAAGGCTAAGCTTCTTCTTTCGGGAGTTAG 901
DB |||||
QY 902 GTCAGGACCATATGATGAATGAAGAAATCTTAA--TTTGGGTCACACCAAGATTGTCTC 960
DB |||||
QY 61 GTCAGGACCATATGATGAATGAAGAAATCTTAAATTTTGGGTCACACCAAGATTGTCTC 120
DB |||||
QY 961 TCTCGAGGTTGGGGGTCCTTAAGTTGGTAGTACCAATACCAATATATACCTAACAA 1020
DB |||||
QY 121 TCTCGAGGTTGGGGGTCCTTAAGTTGGTAGTACCAATACCAATATATACCTAACAA 180
DB |||||
QY 1021 ACCCAATCCATGCTACATACATACATAGCATCCATCCTTGTAGACTGGACCTTCATCA 1080
DB |||||
QY 181 ACCCAATCCATGCTACATACATACATAGCATCCATCCTTGTAGACTGGACCTTCATCA 240
DB |||||

```



```

QY 930 TCTTAATTGGGTCACACCAAGATTGTCTCTCTCGAGTTGGGGGTCCTTAAGTTGG 989
Db 733 TCTTAATTGGGTCACACCAAGATTGTCTCTCTCGAGTTGGGGGTCCTTAAGTTGG 792
QY 990 TAGTAGCAATACCC-AATATATACCTTAACAAACCCATCCATCTACATACATACATAG 1048
Db 793 TAGTAGCAATACCCATTATATATACCTTAACAAACCCATCCATCTACATACATACATAG 852
QY 1049 CATCATCATCTGTAGACTGACCCCTTCATCAAGAGCACCATGGAGGAAGCTTCATCAC 1108
Db 853 CATCATCATCTGTAGACTGACCCCTTCATCAAGAGCACCATGGAGGAAGCTTCATCAC 912
QY 1109 GCCGGCAGCGCATCGCCATTCTCCCACTAGCAGGGCCCTCAAGATACATCGGCTCCT 1168
Db 913 GCCGGCAGCGCATCGCCATTCTCCCACTAGCA-GCCCTCACAAGTACATCGGCTCCT 971
QY 1169 CCTGTTGTCTCTCATGGATCCCTGTCAGAGTTGGAGCTGAGGAGCAGGAAGGCC 1228
Db 972 TCTGGTTGTCTCTCATGGATCCCTGTCAGAGTTGGAGCTGAGGAGCAGGAAGGCC 1031
QY 1229 GAGATCATGGCCAGTCAATCGGTGCAACGGTGGAGCAGCTG 1268
Db 1032 GAGATCATGGCCAGTCAATCGGCGC-ACGGTGGAGCAGCTG 1070

RESULT 4
CC635363/c
LOCUS
DEFINITION
  OGKAW88TV ZM 0.7 1.5 KB Zea mays genomic clone ZMMBma0334P07,
  genomic survey sequence.
ACCESSION
  CC635363
VERSION
  CC635363.1 GI:32013253
KEYWORDS
  GSS.
SOURCE
  Zea mays
  ORGANISM
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
    clade; Panicoideae; Andropogoneae; Zea.
REFERENCE
  1 (bases 1 to 882)
  Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
  Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
  Citek,R.W., Nurnberg,A., Robbins,D. and Lakey,N.
  Consortium for Maize Genomics
  Unpublished (2002)
  Other_GSSs: OGKAW88TH
  Contact: Cathy Whitelaw
  TIGR
  9712 Medical Center Drive, Rockville, MD 20850, USA
  Tel: 301-838-5843
  Fax: 301-838-0208
  Email: whitelaw@tigr.org
  Seq primer: TF
  Class: sheared ends.
FEATURES
  source
    1..882
    /organism="Zea mays"
    /mol_type="genomic DNA"
    /strain="B73"
    /db_xref="taxon:4577"
    /clone="ZMMBma0334P07"
    /clone_lib="ZM 0.7 1.5 KB"
    /note="vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
    methylation filtered genomic DNA library"
ORIGIN
  Query Match 22.1%; Score 860; DB 29; Length 882;
  Best Local Similarity 99.2%; Pred. No. 1.6e-131;
  Matches 875; Conservative 0; Mismatches 5; Indels 2; Gaps 1;
  QY 903 TCAGGGACACATATGAATGAAGAAATCTTAATTTGGGTCACCAAGATTGTCTCTC 962
  Db 882 TCAGGGACACATATGAATGAAGAAATCTTAATTTGGGTCACCAAGATTGTCTCTC 823

```

```

QY 963 TCAGGTTGGGGTCCCTTAAGTTGGTAGCAATACCAATATATACCTAACCAAC 1022
Db 822 TCAGGTTGGGGTCCCTTAAGTTGGTAGCAATACCAATATATACCTAACCAAC 763
QY 1023 CCAATCCATGCTACATACATACATACATACATACATACATACATACATCAAG 1082
Db 762 CCAATCCATGCTACATACATACATACATACATACATACATACATACATCAAG 703
QY 1083 AGCACCATGAGGAAGCTCATCATCAGCCGGCAGCCCATCGCATCTCTCCACATAGCA 1142
Db 702 AGCACCATGAGGAAGCTCATCTCAGCCGGCAGCCCATCGCATCTCTCCACATAGCA 643
QY 1143 GGGCTCACAAGTACATCGGCTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1202
Db 642 GGGCTCACAAGTACATCGGCTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 583
QY 1203 TGGAGCTGAGGAAGCAGAAAGCCCGAGATCATATGGCCAGTCAATCGGTGCAAC 1262
Db 582 TGGAGCTGAGGAAGCAGAAAGCCCGAGATCATATGGCCAGTCAATCGGTGCAAC 523
QY 1263 CAGCTGAGGAATACACACCGGATGCGAGTGGCTGCTCTCTCTCTCTCTCTCTCTCT 1322
Db 522 CAGCTGAGGAATACACACCGGATGCGAGTGGCTGCTCTCTCTCTCTCTCTCTCT 463
QY 1323 ACAGTACCGTGCAGATGCGGTTCACTTCCTACACCTTACATCGCTGACCCGGTGAATGTC 1382
Db 462 ACAGTACCGTGCAGATGCGGTTCACTTCCTACACCTTACATCGCTGACCCGGTGAATGTC 403
QY 1383 GAGCATGTCTCAAGACTAACTTACCAATTACCCAGTAATGACCTGAACTCACTG 1442
Db 402 GAGCATGTCTCAAGACTAACTTACCAATTACCCAGTAATGACCTGAACTCACTG 343
QY 1443 ATGTTGAGTCTTCGGAAATCAGAGCTGAAAGCTGAATCGAATGTGCTGAACACCGTGTA 1502
Db 342 ATGTTGAGTCTTCGGAAATCAGAGCTGAAAGCTGAATCGAATGTGCTGAACACCGTGTA 283
QY 1503 GGAATCGTGTACAGATCTCTACATGGACGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1562
Db 282 GGAATCGTGTACAGATCTCTACATGGACGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 223
QY 1563 CGCGAGCTGTGAGGAAGCAGGAAGAGCGGAGGTTTCAGTTCGCTTCCAGAACCT 1622
Db 222 CGCGAGCTGTGAGGAAGCAGGAAGAGCGGAGGTTTCAGTTCGCTTCCAGAACCT 163
QY 1623 GAGGATTTGAGGCGCATTTGTTTCAGAGAGTACTCCCTGGAAGCTGTCGGGTATCTGAG 1682
Db 162 GAGGATTTGAGGCGCATTTGTTTCAGAGAGTACTCCCTGGAAGCTGTCGGGTATCTGAG 103
QY 1683 CCAGGCATCCAGGCGAGGCAAAAGTTGTGGACATGCAAGGTGAG-ATCACTGCTCCCTTGC 1740
Db 102 CCAGGCATCCAGGCGAGGCAAAAGTTGTGGACATGCAAGGTGAGATATCACTGGTCCCTTGC 43
QY 1741 CATTGCCAATGAGCATTTCAACCTGAGACACGAGAGCTAC 1782
Db 42 CATTGCCAATGAGCATTTCAACCTGAGACACGAGAGCTAC 1

RESULT 5
CG067339/c
LOCUS
DEFINITION
  FUICF65TD ZM 0.6 1.0 KB Zea mays genomic clone ZMMBta0550K10,
  genomic survey sequence.
ACCESSION
  CG067339
VERSION
  CG067339.1 GI:33939519
KEYWORDS
  GSS.
SOURCE
  Zea mays
  ORGANISM
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
    clade; Panicoideae; Andropogoneae; Zea.
  1 (bases 1 to 887)
  Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
REFERENCE
  AUTHORS

```



1313 ACGCCACAGGACAGTACCGCTCGACATCGCCTTCACTTCCCTACACCTACATCGCTGACCC 1372  
Db |  
700 ACGCCACAGGACAGTACCGCTCGACATCGCCTTCACTTCCCTACACCTACATCGCTGACCC 641  
Qy |  
1373 GGTGAATGTTCGAGCATGTCCTCAAGACTTAATCTCACCAGTACCCAGGTAATGACCT 1432  
Db |  
640 GGTGAATGTTCGAGCATGTCCTCAAGACTTAATCTCACCAGTACCCAGGTAATGACCT 581  
Qy |  
1433 GAATCTACTGATGTTCAGTCTTCGGAATCATAGACTGAAGCTGAATGATGTCCTGA 1492  
Db |  
580 GAATCTACTGATGTTCAGTCTTCGGAATCATAGACTGAAGCTGAATGATGTCCTGA 521  
Qy |  
1493 ACACCGTGTAGGAAATCGGTTCACAGATCCTACATGAGAGTCTCTCTGGTGACGGATCT 1552  
Db |  
520 ACACCGTGTAGGAAATCGGTTCACAGATCCTACATGAGAGTCTCTCTGGTGACGGATCT 461  
Qy |  
1553 TCAACGCCGCGAGCGGTGTGGAGGAGACAGAGACGGCGAGTTTCGAGTTTCGCT 1612  
Db |  
460 TCAACGCCGCGAGCGGTGTGGAGGAGACAGAGACGGCGAGTTTCGAGTTTCGCT 401  
Qy |  
1613 CCAAGAACTGAGGATTTACGCGCAATTTGTTTTCAGAGAGTACTTCCCTGAAGCTGTGCG 1672  
Db |  
400 CCAAGAACTGAGGATTTACGCGCAATTTGTTTTCAGAGAGTACTTCCCTGAAGCTGTGCG 341  
Qy |  
1673 GTATCTAGGACGAGCATCCAGGACGACCAAGTTTGGACATGAGAGTACTTCCCTGAAGCTGTGCG 1730  
Db |  
340 GTATCTAGGACGAGCATCCAGGACGACCAAGTTTGGACATGAGAGTACTTCCCTGAAGCTGTGCG 281  
Qy |  
1731 GTCCTCTTGCATTCGCAACATGAGCATTTCAACCTGAGACACGAGAGTACTTCCCTGAAGCTGTGCG 1790  
Db |  
280 GTCCTCTTGCATTCGCAACATGAGCATTTCAACCTGAGACACGAGAGTACTTCCCTGAAGCTGTGCG 221  
Qy |  
1791 TTCAGGAATTTTACATGAGGATGAGTCTCCGAGACAGCTTCCGAGAGTACTTCCCTGAAGCTGTGCG 1850  
Db |  
220 TTCAGGAATTTTACATGAGGATGAGTCTCCGAGACAGCTTCCGAGAGTACTTCCCTGAAGCTGTGCG 161  
Qy |  
1851 AGATCGGACGCTGTCCGCGATCTCCGAGACAGCTTCCGAGAGTACTTCCCTGAAGCTGTGCG 1910  
Db |  
160 AGATCGGACGCTGTCCGCGATCTCCGAGACAGCTTCCGAGAGTACTTCCCTGAAGCTGTGCG 101  
Qy |  
1911 CCAACATCATCATCAGCTGCGGTTCATCGACCGCTGTGGCGCATCAAGAGTTCCTTC 1970  
Db |  
100 CCAACATCATCATCAGCTGCGGTTCATCGACCGCTGTGGCGCATCAAGAGTTCCTTC 41  
Qy |  
1971 ACCTCGGCTCAGAGGCTCTCTAGCGAGAGCATCAAGCT 2010  
Db |  
40 ACCTCGGCTCAGAGGCTCTCTAGCGAGAGCATCAAGCT 1

## RESULT 7

CC685050/c

LOCUS

CG685050 879 bp DNA linear GSS 19-JUN-2003  
CGUFB46TV ZM 0.7-1.5 KB Zea mays genomic clone ZMMBMA0425H20,  
genomic survey sequence.

ACCESSION

CC685050

VERSION

CC685050.1

KEYWORDS

GSS.

SOURCE

Zea mays

ORGANISM

Zea mays

REFERENCE

AUTHORS

Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,

Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfs, T.,

Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.

Consortium for Maize Genomics

Unpublished (2002)

Other\_GSSs: OGUFB46TH

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

FEATURES  
source  
1. 879  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="ZMMBMA0425H20"  
/clone\_lib="ZM 0.7-1.5 KB"  
/notes="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb  
methylation filtered genomic DNA library"

## ORIGIN

Query Match 22.0%; Score 857; DB 29; Length 879;  
Best Local Similarity 99.2%; Pred. No. 5e-131;  
Matches 872; Conservative 0; Mismatches 5; Indels 2; Gaps 1;  
Qy 1228 CGAGATCATGCCAGTCACTCGGTGCAACGGTGGAGCAGCTGAGAACTACCCCGGATGC 1287  
Db |  
879 CGAGATCATGCCAGTCACTCGGTGCAACGGTGGAGCAGCTGAGAACTACCCCGGATGC 820  
Qy |  
1288 ACAGCTGGCTTGTGGGTACCTGTCAACGGCAGGACAGTACCGTTCGACATGCCGTTC 1347  
Db |  
819 ACAGCTGGCTTGTGGGTACCTGTCAACGGCAGGACAGTACCGTTCGACATGCCGTTC 760  
Qy |  
1348 CTTCTCTACCTACATCGCTGACCGGTGAATGTCGAGCATGTCCTCAAGACTTAATCTTC 1407  
Db |  
759 CTTCTCTACCTACATCGCTGACCGGTGAATGTCGAGCATGTCCTCAAGACTTAATCTTC 700  
Qy |  
1408 CCAATTAACCCAAAGTAAATGACCTCACTCACTGATGTTTCACTCTTCGGAATCAGAGC 1467  
Db |  
699 CCAATTAACCCAAAGTAAATGACCTCACTCACTGATGTTTCACTCTTCGGAATCAGAGC 640  
Qy |  
1468 TGAAGCTGAATCGAATGTGCTGAAACACCGTGTAGGAAATCGTGTACAGATCCTACATG 1527  
Db |  
639 TGAAGCTGAATCGAATGTGCTGAAACACCGTGTAGGAAATCGTGTACAGATCCTACATG 580  
Qy |  
1528 GAGCTCTCTCGGTGACGSCATCTTCAACGGCGAGCGGAGCTGTGGAGGAGCAGAGG 1587  
Db |  
579 GAGCTCTCTCGGTGACGSCATCTTCAACGGCGAGCGGAGCTGTGGAGGAGCAGAGG 520  
Qy |  
1588 AAGACGGCGAGTTTCGAGTTTCGCTTCCCAAGAACTTCAAGGAGTTCAGCGCCATTGTGTT 1647  
Db |  
519 AAGACGGCGAGTTTCGAGTTTCGCTTCCCAAGAACTTCAAGGAGTTCAGCGCCATTGTGTT 460  
Qy |  
1648 AGAGATCTCTCTGAGCTGTGGGTATCTAGCCAGGATTCAGGAGGAGGAGGAGGAGGTT 1707  
Db |  
459 AGAGATCTCTCTGAGCTGTGGGTATCTAGCCAGGATTCAGGAGGAGGAGGAGGAGGTT 400  
Qy |  
1708 GTGGACATCAGGTGAG--ATCACTGCTCTCCCTTGCATTCGCAATGCGCAATGAGCATTCACCC 1765  
Db |  
399 GTGGACATCAGGTGAGATATCACTGCTCTCCCTTGCATTCGCAATGAGCATTCACCC 340  
Qy |  
1766 TGAGACACAGAGCTACCTTGGCGATTTCAGGAACCTTTCATGAGGATGACGCTGGACTCC 1825  
Db |  
339 TGAGACACAGAGCTACCTTGGCGATTTCAGGAACCTTTCATGAGGATGACGCTGGACTCC 280  
Qy |  
1826 ATCTGCAAGTTGGGTTCGGGTTCGAGATCGGACATGCTGTCCCGGATCTCCCGGAGAAC 1885  
Db |  
279 ATCTGCAAGTTGGGTTCGGGTTCGAGATCGGACATGCTGTCCCGGATCTCCCGGAGAAC 220  
Qy |  
1886 AGCTTCGCGCAGCGTTTCGATCCGCCACATCATCGTCACGCTCGGTTTCATCCAGCCG 1945  
Db |  
219 AGCTTCGCGCAGCGTTTCGATCCGCCACATCATCGTCACGCTCGGTTTCATCCAGCCG 160  
Qy |  
1946 CTGTGCGCATCAAGAGGTTTCCTTCCACGTCGGGTTCAGAGGCTCTCTAGCGCAGAGCATC 2005  
Db |  
159 CTGTGCGCATCAAGAGGTTTCCTTCCACGTCGGGTTCAGAGGCTCTCTAGCGCAGAGCATC 100  
Qy |  
2006 AAGCTCGTGGACGAGTTTCACTTACAGCGTGTATCCCGGAGGAGGAGGAGGAGGAGTCTCGAG 2065







Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.

REFERENCE  
AUTHORS

1 (bases 1 to 819)  
Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T., Resnick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and Bennetzen, J.

TITLE  
JOURNAL

Maize Genomics Consortium  
Unpublished (2003)

## COMMENT

Other\_GSSs: PUHOF96TD

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TR

Class: sheared ends.

## FEATURES

source

Location/Qualifiers

1..819

/organism="Zea mays"

/mol\_type="genomic DNA"

/strain="B73"

/db\_xref="taxon:4577"

/clone="ZMMBTA496024"

/clone\_lib="ZM\_0.6\_1.0 KB"

/note="Vector: pCR4-ToPO; Site: 1: EcoRI; 0.6-1.0 kb high"

Cor selected genomic DNA library"

## ORIGIN

Query Match 20.4%; Score 793.8; DB 28; Length 819;  
Best Local Similarity 98.9%; Pred. No. 1.3e-120;  
Matches 810; Conservative 0; Mismatches 7; Indels 2; Gaps 1;  
QY 975 GGTCCCTAAGTTGGTAGCAATACCAATATATACCTAACAAACCCATCCATCGT 1034  
DB 1 GATCCCTAAGTTGGTAGCAATACCAATATATACCTAACAAACCCATCCATCGT 60  
QY 1035 ACATACATACATAGCATCCATCACTGTAGACTGACCTTCATCAAGAGCACCATGGAG 1094  
DB 61 ACATACATACATAGCATCCATCACTGTAGACTGACCTTCATCAAGAGCACCATGGAG 120  
QY 1095 GAAGTCTACATCACCGCGGACGCGCATCGCCATTCCTCCACTAGCAGGCGCTCACAG 1154  
DB 121 GAAGTCTACCTCACCGCGGACGCGCATCGCCATTCCTCCACTAGCAGGCGCTCACAG 180  
QY 1155 TACATCGCGCTCCTCCTGGTGTCTCTCATGATCTCTGGTCCAGAGTGGAGCTGAGG 1214  
DB 181 TACATCGCGCTCCTCCTGGTGTCTCTCATGATCTCTGGTCCAGAGTGGAGCTGAGG 240  
QY 1215 AAGCAGAAAGGCGGAGATCATGCGCAGTCTCGGTGCAAGGTTGGAGCAGCTGAGGAC 1274  
DB 241 AAGCAGAAAGGCGGAGATCATGCGCAGTCTCGGTGCAAGGTTGGAGCAGCTGAGGAC 300  
QY 1275 TACACCGGATGACGACTGCTGTGTGGGTACCTGTCCAGGACAGGACAGTCAAGCTC 1334  
DB 301 TACACCGGATGACGACTGCTGTGTGGGTACCTGTCCAGGACAGGACAGTCAAGCTC 360  
QY 1335 GACATCGGTTCACTTCTCACTACCTACATCGCTGACCGGTGAATGCGAGCATGCTC 1394  
DB 361 GACATCGGTTCACTTCTCACTACCTACATCGCTGACCGGTGAATGCGAGCATGCTC 420  
QY 1395 AAGACTAACTTACCAATATCCCAAGGTAAATGCAATGCTGAAACACCGTGTAGGAAATCGTGA 1454  
DB 421 AAGACTAACTTACCAATATCCCAAGGTAAATGCAATGCTGAAATGCAATGCTGAAATCGTGA 480  
QY 1455 CGGAAATCAGAGCTGAAAGCTGAATGCAATGCTGAAATGCTGAAATGCTGAAATCGTGA 1514  
DB 481 CGGAAATCAGAGCTGAAAGCTGAATGCAATGCTGAAATGCTGAAATGCTGAAATCGTGA 540  
QY 1515 CAGATCCTACATGAGCGTGTCTCGGTGACGGCATCTTCAACCGCCGACGCGAGCTGTG 1574  
DB 541 CAGATCCTACATGAGCGTGTCTCGGTGACGGCATCTTCAACCGCCGACGCGAGCTGTG 600

QY 1575 GAGGAAGCAGAGGAAGACGGCGAGTTTCGAGTTCCCTCCCAAGAACCTTGAGGATTTCCAG 1634  
DB 601 GAGGAAGCAGAGGAAGACGGCGAGTTTCGAGTTCCCTCCCAAGAACCTTGAGGATTTCCAG 660  
QY 1635 CCCCATTGTTTCAGAGAGTACTCCCTCAAGCTGTCGGGTATATCTAGCCAGGCAATCCAA 1694  
DB 661 CCCCATTGTTTCAGAGAGTACTCCCTCAAGCTGTCGGGTATATCTAGCCAGGCAATCCAA 720  
QY 1695 GSCAGGCAAGTTGGGACATGCGAGTGTAG--ATCACTGCTCCCTTGCATTCGCAACAT 1752  
DB 721 GSCAGGCAAGTTGGGACATGCGAGTGTAG--ATCACTGCTCCCTTGCATTCGCAACAT 780  
QY 1753 GAGCATTTCAACCTGAGACACGAGAGTACCTTGCCTGA 1791  
DB 781 GAGCATTTCAACCTGAGACACGAGAGTACCTTGCCTGA 819

## RESULT 13

CC723851/c

LOCUS

DEFINITION

CC723851

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Ze mays

Ze mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 848)

White, C.A., Quackenbush, J., Van Aken, S., Utterback, T., Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T., Citek, R.W., Nurnberg, A., Robbins, D. and Lakey, N.

Consortium for Maize Genomics

Unpublished (2002)

Other\_GSSs: OGWER88TH

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TR

Class: sheared ends.

Location/Qualifiers

1..848

/organism="Zea mays"

/mol\_type="genomic DNA"

/strain="B73"

/db\_xref="taxon:4577"

/clone="ZMMBTA0561008"

/clone\_lib="ZM\_0.7\_1.5 KB"

/note="Vector: pBCSK-; Site: 1: HincII; 0.7-1.5 kb"

methylation filtered genomic DNA library"

FEATURES

source

Query Match

Best Local Similarity

Matches

819; Conservative

0; Mismatches

28; Indels

2; Gaps

2;

QY 577 AACAAATTCACCTTTGGTTCACAAATTCCTCTTCAGGATGTACTTTTCACCTTGAACCTGT 636

DB 848 AACAAATTCACCTTTGGTTCACACAAATTCCTCTTCAGGATGTACTTTTCACCTTGAACCT 789

QY 637 CATGTATAGAACAGGAATGGCTCAGTTTTTAAGAAACAATGTAAGATTTTCATTTTCAG 696

DB 788 CATGTATAGAACAGGAATGGCTCAGTTTTTAAGAAACAATGTAAGATTTTCATTTTCAG 729

QY 697 AACTCTTTCT-GGTGGTTGAGTTTCAGACTTTTTCAGACCTGATGATGATGATGATGATGAT 755

DB 728 AACTCTTTCTGGTTGGTTGAGTTTCAGACTTTTTCAGACCTGATGATGATGATGATGATGAT 669

QY 756 TTGTTTCCAAAGTCTGATACAGAACTGGCAACTCCTAATTGATAATA-AAAAGATAA 814  
 Db |||||  
 QY 668 TTGTTTCCAAAGTCTAATAACAGAACTGGCAACTCCTAATTGATAATAAAAAGATAA 609  
 Db |||||  
 QY 815 AATACAGTATCAGATATCTCATTTCTTGGTGGCAGATCACAAGAAAGGAAACACAAAGGC 874  
 Db |||||  
 QY 608 AATACAGTATCAGATATCTCATTTCTTGGTGGCAGATCACAAGAAAGGAAACACAAAGGC 549  
 Db |||||  
 QY 875 TAAAGCCTCTACTTGTTCGGAGTTAGTFCAGGACACCATATGATGAAGAAATCTTA 934  
 Db |||||  
 QY 548 TAAAGCCTCTACTTGTTCGGAGTTAGTFCAGGACACCATATGATGAAGAAATCTTA 489  
 Db |||||  
 QY 935 ATTGGGGTCAACCAAGATTCTCTCTCGAGGTTGGGGGTCCCTAAGGTTGGTAGTA 994  
 Db |||||  
 QY 488 ATTGGGGTCAACCAAGATTCTCTCTCGAGGTTGGGGGTCCCTAAGGTTGGTAGTA 429  
 Db |||||  
 QY 995 GCAATACCAATATATACCTTAAACCAACCAATCCATCTACATACATACATACATCCA 1054  
 Db |||||  
 QY 428 GCAATACCAATATATACCTTAAACCAACCAATCCATCTACATACATACATACATCCA 369  
 Db |||||  
 QY 1055 TCACCTGTAGACTGACCCCTTCATCAAGAGCACCATGGAGAGCTCACATCAGCCGGC 1114  
 Db |||||  
 QY 368 TCACCTGTAGACTGACCCCTTCATCAAGAGCACCATGGAGAGCTCACATCAGCCGGC 309  
 Db |||||  
 QY 1115 GACCCATCGCATCTTCCACTAGCAGGCGCTCACAAAGTACATCGGCTCCTCTGGT 1174  
 Db |||||  
 QY 308 GACCCATCGCATCTTCCACTAGCAGGCGCTCACAAAGTACATCGGCTCCTCTGGT 249  
 Db |||||  
 QY 1175 TGTCTCTCATGATCTGTGTCAGAGTGGAGCTGAGCTGAGGAGCAGAAAGCCGAGATC 1234  
 Db |||||  
 QY 248 TGTCTCTCATGATCTGTGTCAGAGTGGAGCTGAGCTGAGGAGCAGAAAGCCGAGATC 189  
 Db |||||  
 QY 1235 ATGCCAGTATCGGTCAACGGTGGAGAGCTGAGGAGCTGAGGAGCAGAAAGCCGAGATC 1294  
 Db |||||  
 QY 188 ATGCCAGTATCGGTCAACGGTGGAGAGCTGAGGAGCTGAGGAGCAGAAAGCCGAGATC 129  
 Db |||||  
 QY 1295 GCTTGTGGGTACTGTACGGCAGCAGGACAGTGCAGATCGGCTTCACTTCTCTA 1354  
 Db |||||  
 QY 128 GCTTGTGGGTACTGTACGGCAGCAGGACAGTGCAGATCGGCTTCACTTCTCTA 69  
 Db |||||  
 QY 1355 CACCTACATCGTGCACCGGTGATGTCGAGAGTGCCTCAAGACTAACCTACCAATTA 1414  
 Db |||||  
 QY 68 CACCTACATCGTGCACCGGTGATGTCGAGAGTGCCTCAAGACTAACCTACCAATTA 9  
 Db |||||  
 QY 1415 CCCAAGG 1422  
 Db |||||  
 QY 8 CCCAAGG 1

RESULT 14  
 CG280923  
 LOCUS  
 DEFINITION  
 CG280923 811 bp DNA linear GSS 25-AUG-2003  
 genomic survey sequence.  
 CG280923  
 VERSION  
 CG280923.1 GI:34195137  
 KEYWORDS  
 GSS.  
 SOURCE  
 Zea mays  
 ORGANISM  
 Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Zea.  
 1 (bases 1 to 811)  
 Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,  
 Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlking,T.,  
 Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.  
 Consortium for Maize Genomics  
 Unpublished (2002)  
 Other GSSs: OGWLD96TV  
 Contact: Cathy Whitelaw  
 TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA

FEATURES  
 source  
 Location/Qualifiers  
 1..811  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /strain="B73"  
 /db\_xref="taxon:4577"  
 /clone="ZMMBMA0667024"  
 /clone\_lib="ZM 0.7 1.5 KB"  
 /note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb  
 methylation filtered genomic DNA library"  
 ORIGIN  
 Query Match 19.9%; Score 775.2; DB 29; Length 811;  
 Best Local Similarity 98.8%; Pred. No. 1.5e-117;  
 Matches 802; Conservative 0; Mismatches 8; Indels 2; Gaps 2;  
 QY 465 GTCCATCTAATTTTGAAGTGAGGTTCTTAAATTTCAATTTCATTATGCTTCTTTCTAG 524  
 Db 1 GTCCATCTAATTTTGAAGTGAGGTTCTTAAATTTCAATTTCATTATGCTTCTTTCTAG 60  
 QY 525 ACTAGCAACTGCATGACTTTTCACTTTGGGTTCAAAATTTGACTCACAAGAAACAAATT 584  
 Db 61 ACTAGCAATTTGCATGACTTTTCACTTTGGGTTCAAAATTTGACTCACAAGAAACAAATT 120  
 QY 585 CACTTTGGGTTCAAAATTTCTTTCAGGATGACTTTTCACTTGAATGTCATGACTTCAATGATA 644  
 Db 121 CACTTTGGGTTCAAAATTTCTTTCAGGATGACTTTTCACTTGAATGTCATGACTTCAATGATA 179  
 QY 645 GGAACAGGAATGCTCAGTTTAAAGAACATGATGATGATTTCAATTCAGAACTCTTT 704  
 Db 180 GGAACAGGAATGCTCAGTTTAAAGAACATGATGATGATTTCAATTCAGAACTCTTT 239  
 QY 705 CTGGTTGGTGTAGTTTTCAGACTTTTGTACCAAGCTGATGATGATGATTTTCAATTCAGAACTCTTT 764  
 Db 240 CTGGTTGGTGTAGTTTTCAGACTTTTGTACCAAGCTGATGATGATGATTTTCAATTCAGAACTCTTT 299  
 QY 765 AAGTCTGATAACAGAACTGGCAACTCTTAAATGATAATA-AAAAGATAAATAACAGTA 823  
 Db 300 AAGTCTTAATACAGAACTGGCAACTCTTAAATGATAATAATAAATAACAGTA 359  
 QY 824 TCAGATATCTCATTTTCTTGTTCAGATCACAAGAAAGAACACAAGGCTAAGCTCC 883  
 Db 360 TCAGATATCTCATTTTCTTGTTCAGATCACAAGAAAGAACACAAGGCTAAGCTCC 419  
 QY 884 TACTTGTTCGGAGTTAGTCTCAGGACACCATATGAATGAAAGAAATCTTAATTTGGGT 943  
 Db 420 TACTTGTTCGGAGTTAGTCTCAGGACACCATATGAATGAAAGAAATCTTAATTTGGGT 479  
 QY 944 CACACCAAGATTGCTCTCTGAGTTGGGGGTCCCTAAGGTTGGTGTAGTACCAATACCC 1003  
 Db 480 CACACCAAGATTGCTCTCTGAGTTGGGGGTCCCTAAGGTTGGTGTAGTACCAATACCC 539  
 QY 1004 AATATATCACTTAACAAACCAATCCATGCTACATACATACATAGCATCCATCTGTA 1063  
 Db 540 AATATATCACTTAACAAACCAATCCATGCTACATACATACATAGCATCCATCTGTA 599  
 QY 1064 GACTGGACCTTTCATCAAGAGCACCATGAGGAGCTCACATCAGCGGGGAGCGCATC 1123  
 Db 600 GACTGGACCTTTCATCAAGAGCACCATGAGGAGCTCACATCAGCGGGGAGCGCATC 659  
 QY 1124 GCCATTCTTCCACTAGCAGGCTCTCAAGATAATCGCGCTCTCTCTGTTGCTCTC 1183  
 Db 660 GCCATTCTTCCACTAGCAGGCTCTCAAGATAATCGCGCTCTCTCTGTTGCTCTC 719  
 QY 1184 ATGGATCTGCTCCAGAGTGGAGCTTGAGGAGCAGAAAGCCGAGATCATGGCAGT 1243  
 Db 720 ATGGATCTGCTCCAGAGTGGAGCTTGAGGAGCAGAAAGCCGAGATCATGGCAGT 779

QY 756 TTGTTTCCAAAGTCTGATACAGAACTGGCAACTCCTAATTGATAATA-AAAAGATAA 814  
 Db |||||  
 QY 668 TTGTTTCCAAAGTCTAATAACAGAACTGGCAACTCCTAATTGATAATAAAAAGATAA 609  
 Db |||||  
 QY 815 AATACAGTATCAGATATCTCATTTCTTGGTGGCAGATCACAAGAAAGGAAACACAAAGGC 874  
 Db |||||  
 QY 608 AATACAGTATCAGATATCTCATTTCTTGGTGGCAGATCACAAGAAAGGAAACACAAAGGC 549  
 Db |||||  
 QY 875 TAAAGCCTCTACTTGTTCGGAGTTAGTFCAGGACACCATATGATGAAGAAATCTTA 934  
 Db |||||  
 QY 548 TAAAGCCTCTACTTGTTCGGAGTTAGTFCAGGACACCATATGATGAAGAAATCTTA 489  
 Db |||||  
 QY 935 ATTGGGGTCAACCAAGATTCTCTCTCGAGGTTGGGGGTCCCTAAGGTTGGTAGTA 994  
 Db |||||  
 QY 488 ATTGGGGTCAACCAAGATTCTCTCTCGAGGTTGGGGGTCCCTAAGGTTGGTAGTA 429  
 Db |||||  
 QY 995 GCAATACCAATATATACCTTAAACCAACCAATCCATCTACATACATACATACATCCA 1054  
 Db |||||  
 QY 428 GCAATACCAATATATACCTTAAACCAACCAATCCATCTACATACATACATACATCCA 369  
 Db |||||  
 QY 1055 TCACCTGTAGACTGACCCCTTCATCAAGAGCACCATGGAGAGCTCACATCAGCCGGC 1114  
 Db |||||  
 QY 368 TCACCTGTAGACTGACCCCTTCATCAAGAGCACCATGGAGAGCTCACATCAGCCGGC 309  
 Db |||||  
 QY 1115 GACCCATCGCATCTTCCACTAGCAGGCGCTCACAAAGTACATCGGCTCCTCTGGT 1174  
 Db |||||  
 QY 308 GACCCATCGCATCTTCCACTAGCAGGCGCTCACAAAGTACATCGGCTCCTCTGGT 249  
 Db |||||  
 QY 1175 TGTCTCTCATGATCTGTGTCAGAGTGGAGCTGAGCTGAGGAGCAGAAAGCCGAGATC 1234  
 Db |||||  
 QY 248 TGTCTCTCATGATCTGTGTCAGAGTGGAGCTGAGCTGAGGAGCAGAAAGCCGAGATC 189  
 Db |||||  
 QY 1235 ATGCCAGTATCGGTCAACGGTGGAGAGCTGAGGAGCTGAGGAGCAGAAAGCCGAGATC 1294  
 Db |||||  
 QY 188 ATGCCAGTATCGGTCAACGGTGGAGAGCTGAGGAGCTGAGGAGCAGAAAGCCGAGATC 129  
 Db |||||  
 QY 1295 GCTTGTGGGTACTGTACGGCAGCAGGACAGTGCAGATCGGCTTCACTTCTCTA 1354  
 Db |||||  
 QY 128 GCTTGTGGGTACTGTACGGCAGCAGGACAGTGCAGATCGGCTTCACTTCTCTA 69  
 Db |||||  
 QY 1355 CACCTACATCGTGCACCGGTGATGTCGAGAGTGCCTCAAGACTAACCTACCAATTA 1414  
 Db |||||  
 QY 68 CACCTACATCGTGCACCGGTGATGTCGAGAGTGCCTCAAGACTAACCTACCAATTA 9  
 Db |||||  
 QY 1415 CCCAAGG 1422  
 Db |||||  
 QY 8 CCCAAGG 1

QY	1244	CATCGGTGCAACGGTGGAGCAGCTGAGGAAC	1275
Db	780	CATCGGCGCAACGGTGGAGCAGCTGAGGAAC	811
RESULT 15	AY110120/c		
DEFINITION	Zea mays CL58113_1 mRNA sequence.	997 bp	linear
ACCESSION	AY110120		
VERSION	AY110120.1	GI:21214224	
KEYWORDS	HTC		
SOURCE	Zea mays		
ORGANISM	Zea mays		
REFERENCE	Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD		
AUTHORS	1 (bases 1 to 997)		
TITLE	Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whitsitt, M.S., Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.		
JOURNAL	Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes		
REFERENCE	Unpublished (2002)		
AUTHORS	2 (bases 1 to 997)		
TITLE	Coe, E.H.		
JOURNAL	Direct Submission		
COMMENT	Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA		
FEATURES	Location/Qualifiers		
source	1..997		
	/organism="Zea mays"		
	/mol_type="mRNA"		
	/db_xref="MaizeDB:632508"		
	/db_xref="taxon:4577"		
	/clone_lib="Maize Mapping Project/DuPont Consensus Library"		
	/note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed Dupont contigs; this resource was assembled by Dupont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"		
ORIGIN			
Query Match	19.9%	Score 774.2;	DB 11; Length 997;
Best Local Similarity	81.6%	Pred. No. 2.3e-117;	
Matches 888;	Conservative	0; Mismatches 108;	Indels 92; Gaps 2;
QY	2194	AAATGGGCATGAGCAGACACTCTGTGTCACGGTTTCATCGAGCTAGCGCGCGG	2253
Db	997	AAAGGGGCATGAGCAGACACTCTGTGTCACGGTTTCATCGAGCTAGCGCGCGG	2253